B The Saas Fee Lectures on Strong Gravitational Lensing

C.S. Kochanek

Department of Astronomy, The Ohio State University
email: ckochanek@astronomy.ohio-state.edu

B.1 Introduction

The objective of these lectures is to provide a practical introduction to strong gravitational lensing including the data, the theory, and the application of strong lensing to other areas of astrophysics. This is Part 2 of the complete Saas Fee lectures on gravitational lensing. Part 1 (Schneider 2004) provides a basic introduction, Part 2 (Kochanek 2004) examines strong gravitational lenses, Part 3 (Schneider 2004) explores cluster lensing and weak lensing, and Part 4 (Wambsganss 2004) examines microlensing.[1] The complete lectures provide an updated summary of the field from Schneider, Ehlers & Falco (1992). There are also many earlier (and shorter!) reviews of strong lensing (e.g. Blandford & Kochanek 1987a, Blandford & Narayan 1992, Refsdal & Surdej 1994, Wambsganss 1998, Narayan & Bartelmann 1999, Courbin, Saha & Schechter 2002, Claeskens & Surdej 2002).

It is not my objective in this lecture to provide a historical review, carefully outlining the genealogy of every development in gravitational lensing, but to focus on current research topics. Part 1 of these lectures summarizes the history of lensing and introduces most of the basic equations of lensing. The present discussion is divided into 9 sections. We start in §B.2 with an introduction to the observational data. In §B.3 we outline the basic principles of strong lenses, building on the general theory of lensing from Part 1. In §B.4 we discuss modeling gravitational lenses and the determination of the mass distribution of lens galaxies. In §B.5 we discuss time delays and the Hubble constant. In §B.6 we discuss gravitational lens statistics and the cosmological model. In §B.7 we discuss the differences between galaxies and clusters as lenses. In §B.8 we discuss the effects of substructure or satellites on gravitational lenses. In §B.9 we discuss the optical properties of lens galaxies and in §B.10 we discuss extended sources and quasar host galaxies. Finally in §B.11 we discuss the future of strong gravitational lensing.

It will be clear to readers already familiar with the field that these are my lectures on strong lensing rather than an attempt to achieve a mythical

[1] For astro-ph users, the lectures should be referenced as: Kochanek, C.S., Schneider, P., Wambsganss, J., 2004, Gravitational Lensing: Strong, Weak & Micro, Proceedings of the 33rd Saas-Fee Advanced Course, G. Meylan, P. Jetzer & P. North, eds. (Springer-Verlag: Berlin).

consensus. I have tried to make clear what matters (and what does not), what lensing can do (and cannot do) for astrophysics, where the field is serving the community well (and poorly), and where non-experts have understood the consequences (or have failed to do so). Doing so requires having definite opinions with which other researchers may well disagree. We will operate on the assumption that anyone who disagrees sufficiently violently will have an opportunity to wreak a horrible revenge at a later date by spending six months doing their own set of lectures.

- B.1 Introduction 1
- B.2 An Introduction to the Data 2
- B.3 Basic Principles 9
  1. Some Nomenclature 9
  2. Circular Lenses 11
  3. Non-Circular Lenses 23
- B.4 The Mass Distributions of Galaxies 33
  1. Common Models for the Monopole 37
  2. The Effective Single Screen Lens 41
  3. Constraining the Monopole 42
  4. The Angular Structure of Lenses 48
  5. Constraining Angular Structure 52
  6. Model Fitting and the Mass Distribution of Lenses 55
  7. Non-Parametric Models 61
  8. Statistical Constraints on Mass Distributions 64
  9. Stellar Dynamics and Lensing 71
- B.5 Time Delays 76
  1. A General Theory of Time Delays 78
  2. Time Delay Lenses in Groups or Clusters 81
  3. Observing Time Delays and Time Delay Lenses 83
  4. Results: The Hubble Constant and Dark Matter 87
  5. The Future of Time Delay Measurements 94
- B.6 Gravitational Lens Statistics 95
  1. The Mechanics of Surveys 95
  2. The Lens Population 99
  3. Cross Sections 105
  4. Optical Depth 107
  5. Spiral Galaxy Lenses 109
  6. Magnification Bias 110
  7. Cosmology With Lens Statistics 118
  8. The Current State 119
- B.7 What Happened to The Cluster Lenses? 123
  1. The Effects of Halo Structure and the Power Spectrum 131
  2. Binary Quasars 132
- B.8 The Role of Substructure 134
  1. Low Mass Dark Halos 144

- B.9 The Optical Properties of Lens Galaxies 146
  1. The Interstellar Medium of Lens Galaxies 152
- B.10 Extended Sources and Quasar Host Galaxies 157
  1. An Analytic Model for Einstein Rings 158
  2. Numerical Models of Extended Lensed Sources 162
  3. Lensed Quasar Host Galaxies 166
- B.11 Does Strong Lensing Have A Future? 169

B.2 An Introduction to the Data

There are now 82 candidates for multiple image lenses besides those found in rich clusters. Of these candidates, there is little doubt about 74 of them. The ambiguous candidates consist of faint galaxies with nearby arcs and no spectroscopic data. Indeed, the absence of complete spectroscopic information is the bane of most astrophysical applications of lenses. Less than half (38) of the good candidates have both source and lens redshifts – 43 have lens redshifts, 64 have source redshifts, and 5 have neither redshift. Much of this problem could be eliminated in about 5 clear nights of 8m time, but no TAC seems willing to devote the effort even though lens redshifts probably provide more cosmological information per redshift than any other sparsely distributed source. Of these 74 lenses, 10 have had their central velocity dispersions measured and 10 have measured time delays. A reasonably complete summary of the lens data is available at the CASTLES WWW site http://cfa-www.harvard.edu/castles/, although lack of manpower means that it is updated only episodically.

Fig. B.1 shows the distribution of the lenses in image separation and source redshift. The separations of the images range from $0\farcs35$ to $15\farcs9$ (using either half the image separations or the mean distance of the images from the lens). The observed distribution combines both the true separation distribution and selection effects. For example, in simple statistical models using standard models for galaxy properties we would expect to find that the logarithmic separation distribution $dN/d\ln\Delta\theta$ is nearly constant at small separations (i.e. $dN/d\Delta\theta \propto \Delta\theta$, §B.6), while the raw, observed distribution shows a cutoff due to the finite resolution of lens surveys (typically $0\farcs25$ to $1\farcs0$ depending on the survey). The cutoff at larger separations is real, and it is a consequence of the vastly higher lensing efficiency of galaxies relative to clusters created by the cooling of the baryons in galaxies (see §B.7).

Fig. B.2 shows the distribution in image separation and lens galaxy redshift. There is no obvious trend in the typical separation with redshift, as might be expected if there were rapid evolution in the typical masses of galaxies. Unfortunately, there is also an observational bias to measure the redshifts of large separation lenses, where the lens galaxies tend to be brighter and less confused with the images, which makes quantitative interpretation of any trends in separation with redshift difficult. There is probably also a

bias against finding large separation, low lens redshift systems because the flux from the lens galaxy will more easily mask the flux from the source. We examine the correlations between image separations and lens magnitudes in §B.9.

In almost all cases the lenses have geometries that are "standard" for models in which the angular structure of the gravitational potential is dominated by the quadrupole moments of the density distribution, either because the lens is ellipsoidal or because the lens sits in a strong external (tidal) shear field. Of the 60 lenses where a compact component (quasar or radio core) is clearly identifiable, 36 are doubles, 2 are triples, 20 are quads, 1 has five images and 1 has six images. The doubles and quads are the standard geometries produced by standard lenses with nearly singular central surface densities. Examples of these basic patterns are shown in Figs. B.3 and B.4.

In a two-image lens like HE1104–1805 (Wisotzki et al. 1993), the images usually lie at markedly different distances from the lens galaxy because the source must be offset from the lens center to avoid producing four images. The quads show three generic patterns depending on the location of the source relative to the lens center and the caustics. There are cruciform quads like HE0435–1223 (Wisotzki et al. 2002), where the images form a cross pattern bracketing the lens. These are created when the source lies almost directly behind the lens. There are fold-dominated quads like PG1115+080 (Weymann et al. 1980), where the source is close to a fold caustic and we observe a close pair of highly magnified images. Finally, there are cusp-dominated quads like RXJ1131–1231 (Sluse et al. 2003), where the source is close to a cusp caustic and we observe a close triple of highly magnified images. These are all generic patterns expected from caustic theory, as we discuss in Part 1 and §B.3. We will discuss the relative numbers of doubles and quads in §B.6.

The lenses with non-standard geometries all have differing origins. One triple, APM08279+5255 (Irwin et al. 1998, Ibata et al. 1999, Muñoz, Kochanek & Keeton 2001), is probably an example of a disk or exposed cusp lens (see §B.3), while the other, PMNJ1632–0033 (Winn et al. 2002a, Winn, Rusin & Kochanek 2004), appears to be a classical three image lens with the third image in the core of the lens (Fig. B.5). The system with five images, PMNJ0134–0931 (Winn et al. 2002c, Keeton & Winn 2003, Winn et al. 2003), is due to having two lens galaxies, while the system with six images, B1359+154 (Myers et al. 1999, Rusin et al. 2001), is a consequence of having three lens galaxies inside the Einstein ring. Many lenses have luminous satellites that are required in any successful lens model, such as the satellites known as "Object X" in MG0414+0534 (Hewitt et al. 1992, Schechter & Moore 1993) and object D in MG2016+112 (Lawrence et al. 1984) shown in Fig. B.6. These satellite galaxies can be crucial parts of lens models, although there has been no systematic study of their properties in the lens sample.

If the structure of the source is more complicated, then the resulting image geometries become more complicated. For example, the source of the

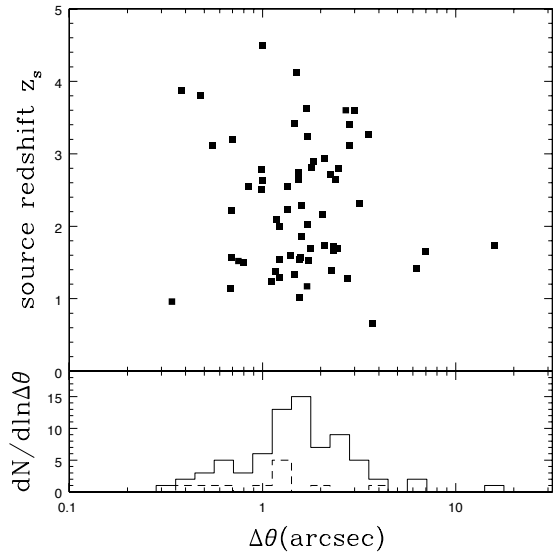


Fig. B.1. The distribution of lens galaxies in separation $\Delta\theta$ and source redshift z_s. The solid histogram shows the distribution in separation for all the lenses while the dashed histogram shows the distribution of those with unmeasured source redshifts.

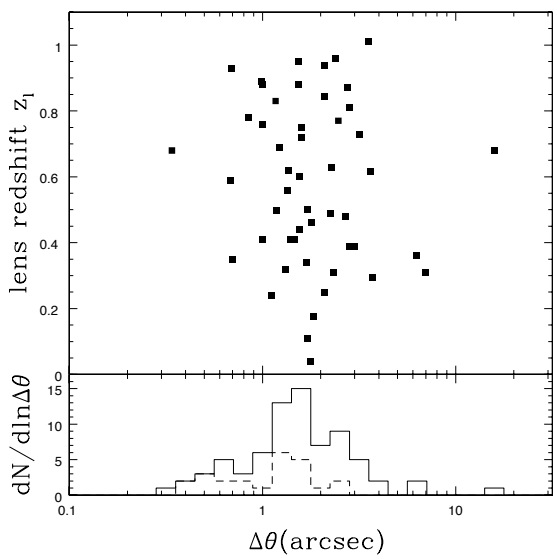


Fig. B.2. The distribution of lens galaxies in separation $\Delta\theta$ and lens redshift z_l. The solid histogram shows the distribution in separation for all the lenses while the dashed histogram shows the distribution of those with unmeasured lens redshifts. There are no obvious correlations between lens redshift z_l and separation $\Delta\theta$, but the strong selection bias that small separation lenses are less likely to have measured redshifts makes this difficult to interpret. There may also be a deficit of low redshift, large separation lenses, which may be a selection bias created by the difficulty of finding quasar lenses embedded in bright galaxies.

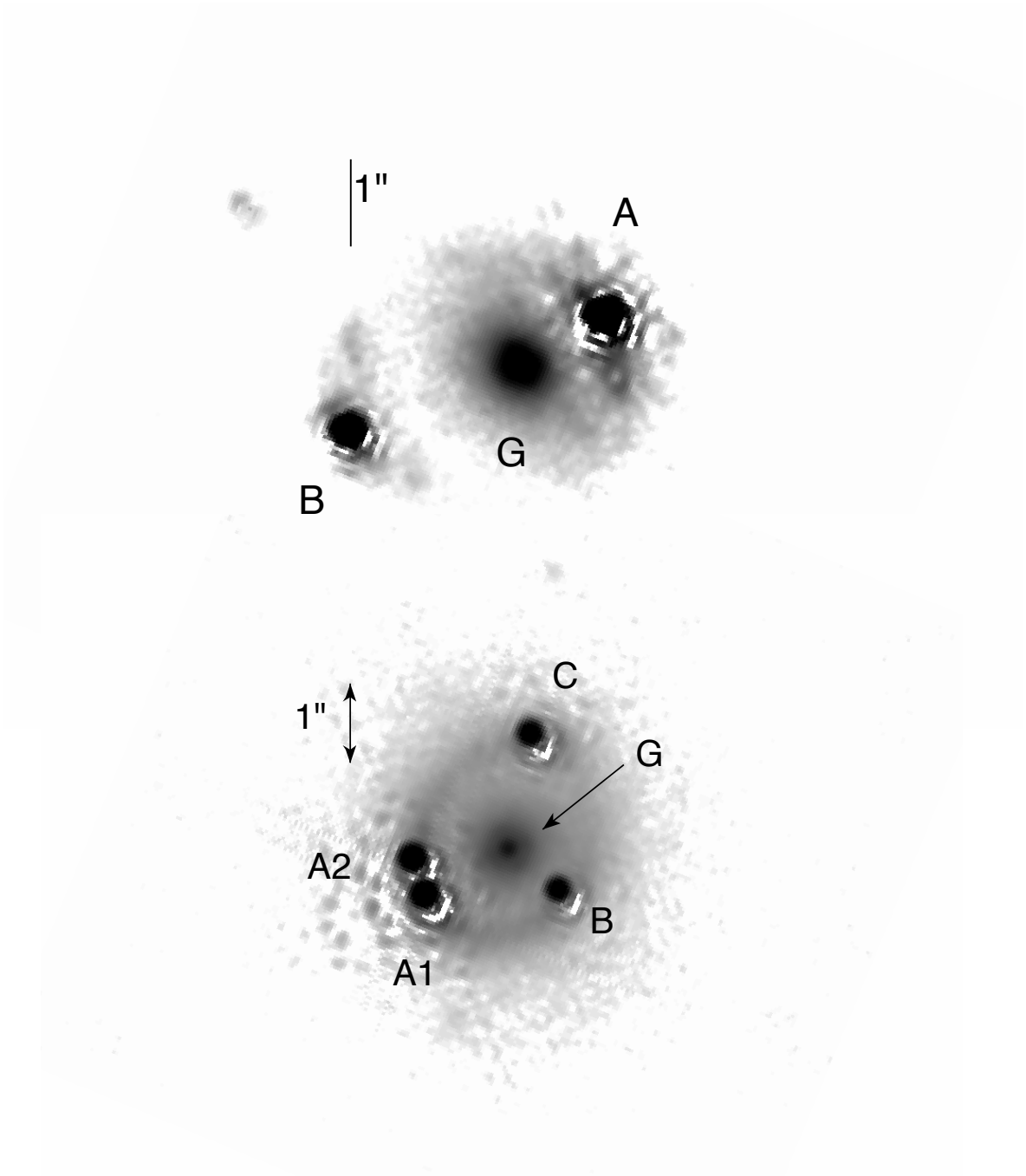
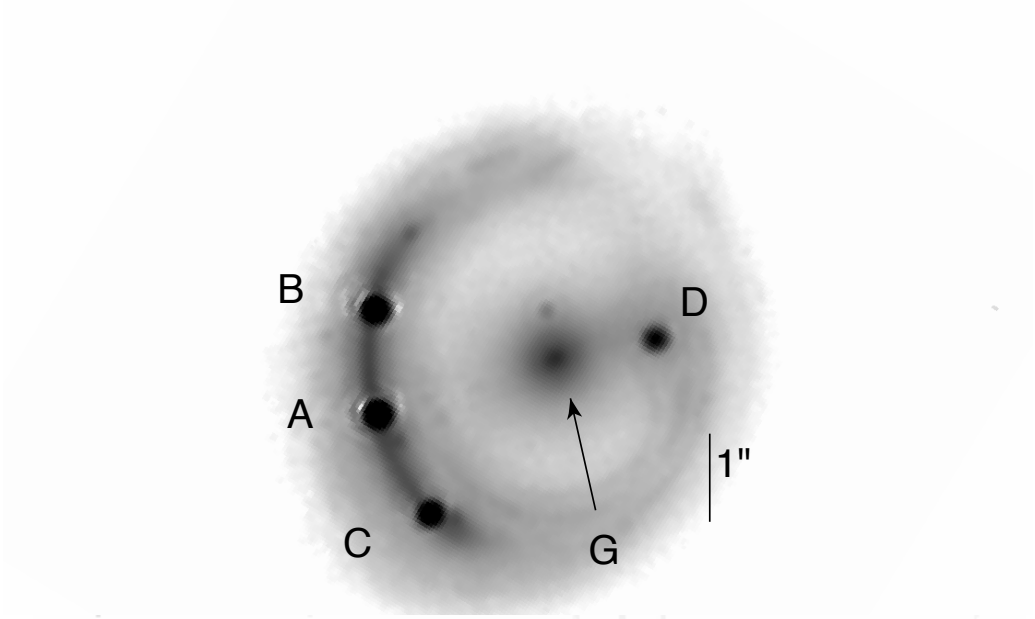


Fig. B.3. Standard image geometries. (Top) The two-image lens HE1104–1805. G is the lens galaxy and A and B are the quasar images. We also see arc images of the quasar host galaxy underneath the quasar images. (Bottom) The four-image lens PG1115+080 showing the bright A_1 and A_2 images created by a fold caustic. (Top, next page) The four-image lens RXJ1131–1231 showing the bright A, B and C images created by a cusp caustic. (Bottom, next page) The four-image lens HE0435–1223, showing the cruciform geometry created by a source near the center of the lens. For each lens we took the CASTLES H-band image, subtracted the bright quasars and then added them back as Gaussians with roughly the same FWHM as the real PSF. This removes the complex diffraction pattern of the HST PSF and makes it easier to see low surface brightness features.



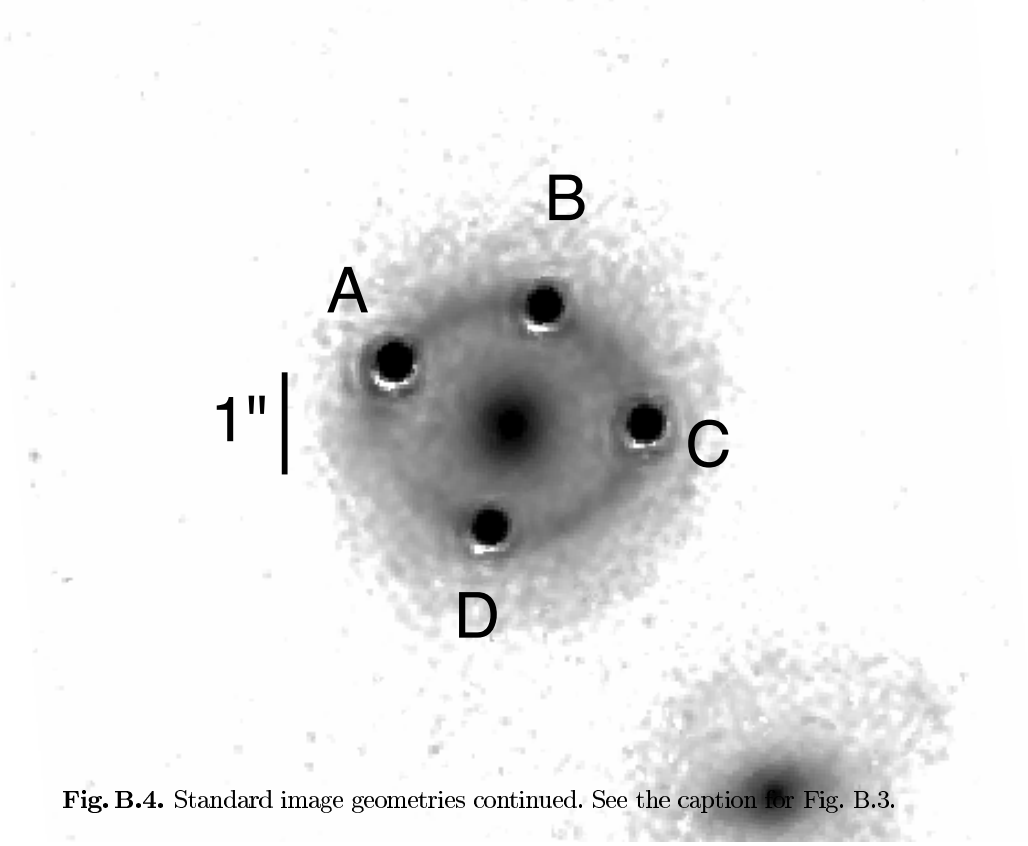


Fig. B.4. Standard image geometries continued. See the caption for Fig. B.3.

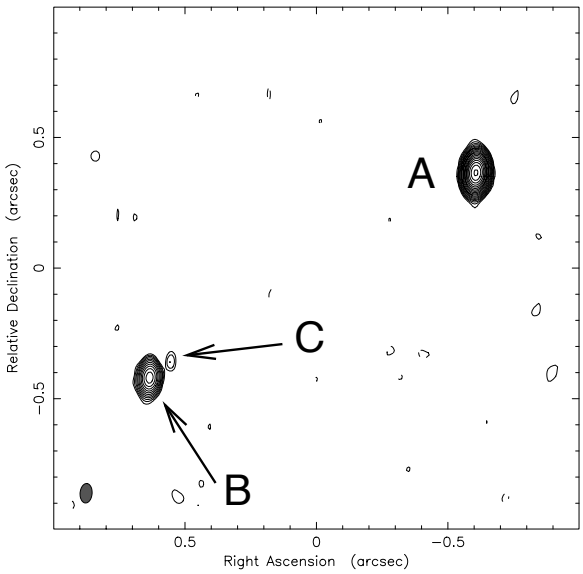


Fig. B.5. PMN1632–0033 is the only known lens with a "classical" third image in the core of the lens galaxy. The center of the lens galaxy is close to the faint C image. Images A, B and C have identical radio spectra except for the longest wavelength flux of C, which can be explained by absorption in the core of the lens galaxy. Time delay measurements would be required to make the case absolutely secure. A central black hole in the lens galaxy might produce an additional image with a flux about 10% that of image C. (Winn et al. 2004)

radio lens B1933+503 (Sykes et al. 1998) consists of a radio core and two radio lobes, leading to 10 observable images because the core and one lobe are quadruply imaged and the other lobe is doubly imaged (Fig. B.7). If instead of discrete emission peaks there is a continuous surface brightness distribution, then we observe arcs or rings surrounding the lens galaxy. Fig. B.8 shows examples of Einstein rings for the case of MG1131+0456 (Hewitt et al. 1988) in both the radio (Chen & Hewitt 1993) and the infrared (Kochanek et al. 2000). The radio ring is formed from an extended radio jet, while the infared ring is formed from the host galaxy of the radio source. We also chose most of the examples in Figs. B.3 and B.4 to show prominent arcs and rings formed by lensing the host galaxy of the source quasar. We discuss arcs and rings in §B.10.

B.3 Basic Principles

Most gravitational lenses have the standard configurations we illustrated in §B.2. These configurations are easily understood in terms of the caustic struc-

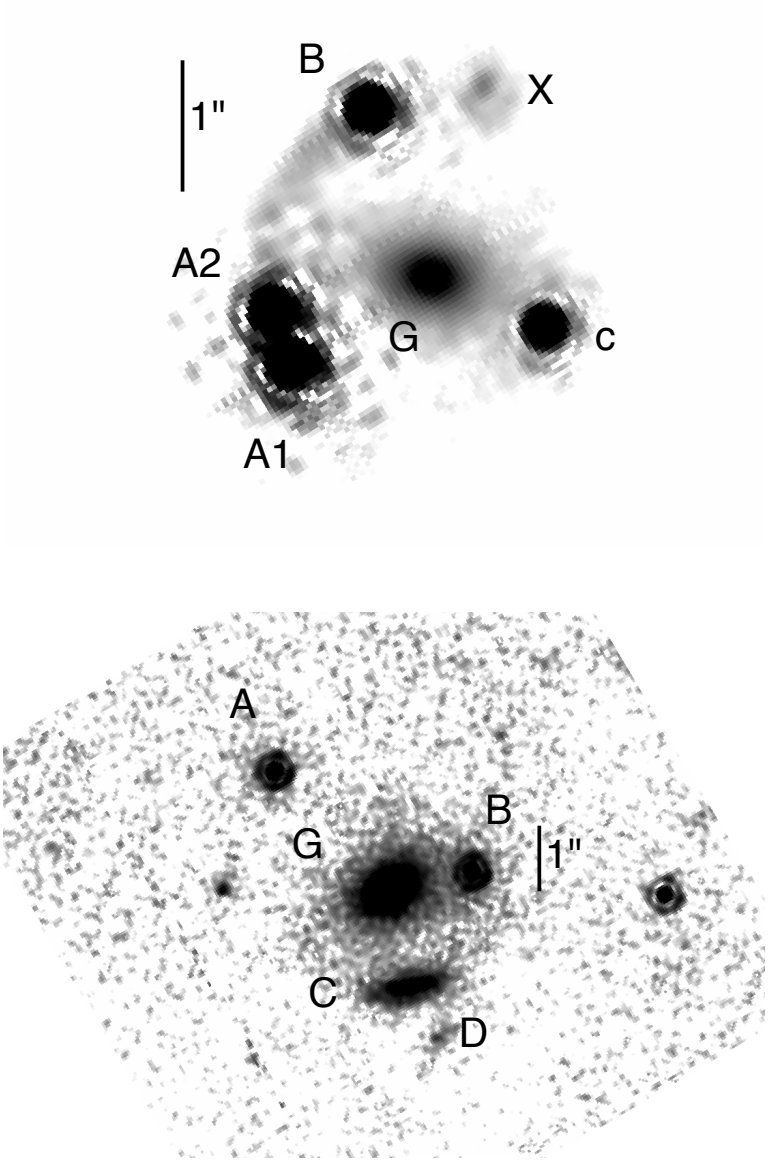


Fig. B.6. H-band images of two lenses with small companions that are crucial for successful models. The upper image shows "Object X" in MG0414+0534, and the lower image shows component D of MG2016+112. MG2016+112 has the additional confusion that only A and B are images of the quasar (Koopmans et al. 2002). Image C is some combination of emission from the quasar jet (it is an extended X-ray source, Chartas et al. 2001) and the quasar host galaxy. Object D is known to be at the same redshift as the primary lens galaxy G (Koopmans & Treu 2002).

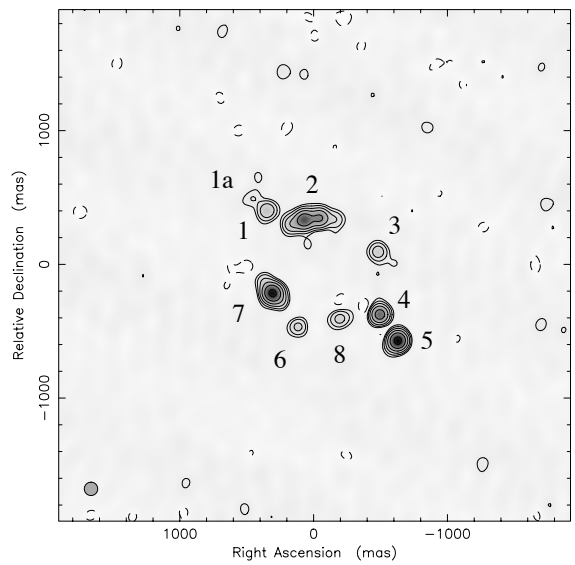


Fig. B.7. A Merlin map of B1933+503 showing the 10 observed images of the three component source (Marlow et al. 1999). The flat radio spectrum core is lensed into images 1, 3, 4 and 6. One radio lobe is lensed into images 1a and 8, while the other is lensed into images 2, 7 and 5. Image 2 is really two images merging on a fold.

tures generic to simple lens models. In this section we illustrate the origin of these basic geometries using simple mathematical examples. We build on the general outline of lensing theory from Part 1.

B.3.1 Some Nomenclature

Throughout this lecture we use comoving angular diameter distances (also known as proper motion distances) rather than the more familiar angular diameter distances because almost every equation in gravitational lensing becomes simpler. The distance between two redshifts i and j is

$$D_{ij} = \frac{r_H}{|\Omega_k|^{1/2}} \mathrm{sinn}\left[\int_i^j \frac{|\Omega_k|^{1/2} dz}{\left[(1+z)^2(1+\Omega_M z) - z(2+z)\Omega_\Lambda\right]^{1/2}}\right] \tag{B.1}$$

where Ω_M, Ω_Λ and $\Omega_k = 1 - \Omega_M - \Omega_\Lambda$ are the present day matter density, cosmological constant and "curvature" density respectively, $r_H = c/H_0$ is the Hubble radius, and the function $\mathrm{sinn}(x)$ becomes $\sinh(x)$, x or $\sin(x)$ for open ($\Omega_k > 0$), flat ($\Omega_k = 0$) and closed ($\Omega_k < 0$) models (Carroll,

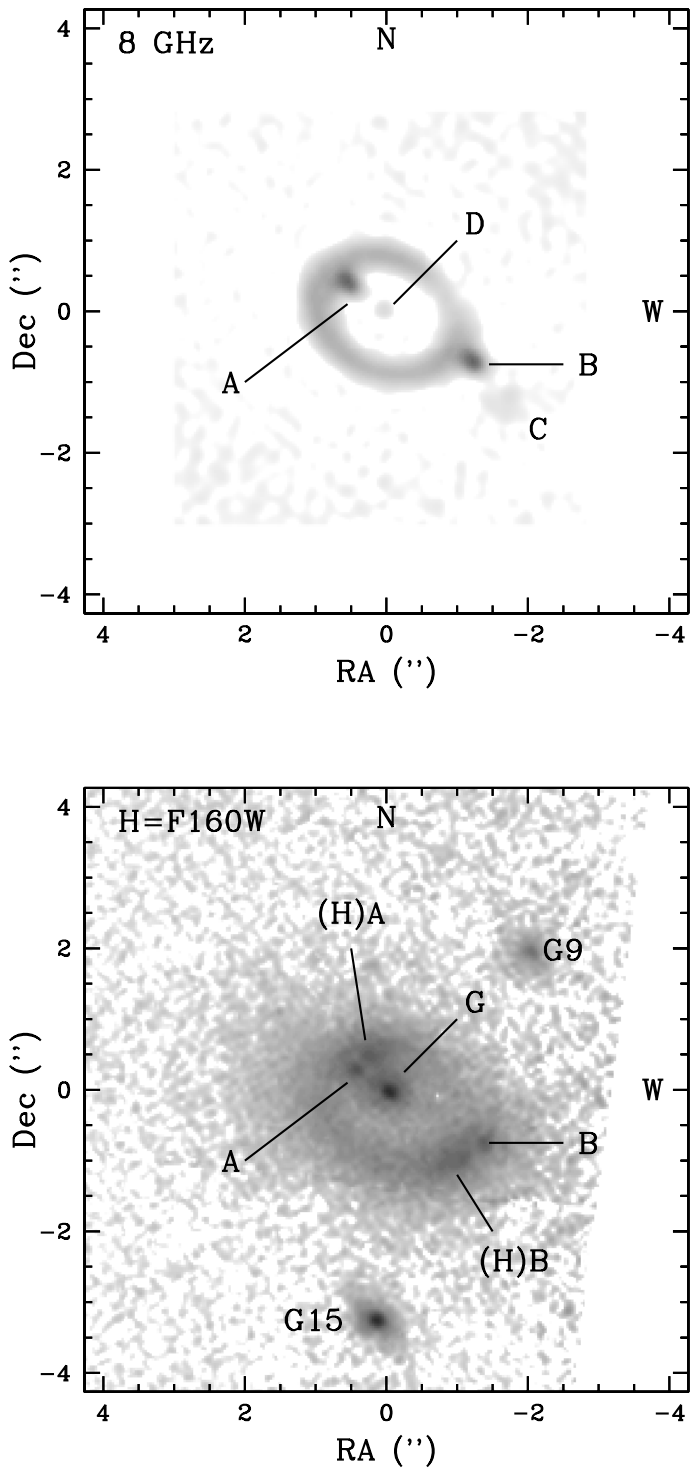


Fig. B.8. The radio (top) and H-band (bottom) rings in MG11131+0456. The radio map was made at 8 GHz by Chen & Hewitt (1993), while the H-band image is from Kochanek et al. (2000). The radio source D is probably another example of a central odd image, but the evidence is not as firm as that for PMN1632−0033. Note the perturbing satellite galaxies (G9 and G15) in the H-band image.

Press & Turner 1992). We use D_d, D_s and D_{ds} for the distances from the observer to the lens, from the observer to the source and from the lens to the source. These distances are trivially related to the angular diameter distances, $D_{ij}^{ang} = D_{ij}/(1+z_j)$, and luminosity distances, $D_{ij}^{lum} = D_{ij}(1+z_j)$. In a flat universe, one can simply add comoving angular diameter distances ($D_s = D_d + D_{ds}$), which is not true of angular diameter distances. The comoving volume element is

$$dV = \frac{4\pi D_d^2 dD_d}{\left(1+\Omega_k r_H^{-2} D_d^2\right)^{1/2}} \rightarrow 4\pi D_d^2 dD_d \tag{B.2}$$

for flat universes. We denote angles on the lens plane by $\boldsymbol{\theta} = \theta(\cos\chi, \sin\chi)$ and angles on the source plane by $\boldsymbol{\beta}$. Physical lengths on the lens plane are $\boldsymbol{\xi} = D_d^{ang}\boldsymbol{\theta}$. The lensing potential, denoted by $\Psi(\boldsymbol{\theta})$, satisfies the Poisson equation $\nabla^2\Psi = 2\kappa$ where $\kappa = \Sigma/\Sigma_c$ is the surface density Σ in units of the critical surface density $\Sigma_c = c^2(1+z_l)D_s/(4\pi G D_d D_{ds})$. For a more detailed review of the basic physics, see Part 1.

B.3.2 Circular Lenses

While one of the most important lessons about modeling gravitational lenses in the real world is that you can never (EVER!)[2] safely neglect the angular structure of the gravitational potential, it is still worth starting with circular lens models. They provide a basic introduction to many of the elements which are essential to realistic models without the need for numerical calculation. In a circular lens, the effective lens potential (Part 1) is a function only of the distance from the lens center $\theta = |\boldsymbol{\theta}|$. Rays are radially deflected by the angle

$$\alpha(\theta) = \frac{2}{\theta}\int_0^\theta \theta d\theta \kappa(\theta) = \frac{4GM(<\xi)}{c^2\xi}\frac{D_{ds}}{D_s} \tag{B.3}$$

where we recall from Part 1 that $\kappa(\theta) = \Sigma(\theta)/\Sigma_c$ is the surface density in units of the critical surface density, D_{ds} and D_s are the lens-source and observer-source comoving distances and $\xi = D_d^{ang}\theta$ is the proper distance from the lens. The bend angle is simply twice the Schwarzschild radius of the enclosed mass, $4GM(<\xi)/c^2$, divided by the impact parameter ξ and scaled by the distance ratio D_{ds}/D_s.

The lens equation (see Part 1) becomes

$$\boldsymbol{\beta} = \boldsymbol{\theta}\left[1 - \alpha(\theta)/\theta\right] = \boldsymbol{\theta}\left[1 - \langle\kappa(\theta)\rangle\right] \tag{B.4}$$

where

$$\langle\kappa(\theta)\rangle = \frac{2}{\theta^2}\int_0^\theta \theta d\theta \kappa(\theta) = \alpha(\theta)/\theta \tag{B.5}$$

[2] AND I MEAN EVER! DON'T EVEN THINK OF IT!

is the average surface density interior to θ in units of the critical density. Note that there must be a region with $\langle\kappa\rangle > 1$ to have solutions on both sides of the lens center. Because of the circular symmetry, all images will lie on a line passing through the source and the lens center.

The inverse magnification tensor (or Hessian, see Part 1) also has a simple form, with

$$M^{-1} = \frac{d\boldsymbol{\beta}}{d\boldsymbol{\theta}} = (1-\kappa)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \gamma \begin{pmatrix} \cos 2\chi & \sin 2\chi \\ \sin 2\chi & -\cos 2\chi \end{pmatrix} \tag{B.6}$$

where $\boldsymbol{\theta} = \theta(\cos\chi, \sin\chi)$. The convergence (surface density) is

$$\kappa = \frac{1}{2}\left(\frac{\alpha}{\theta} + \frac{d\alpha}{d\theta}\right) \tag{B.7}$$

and the shear is

$$\gamma = \frac{1}{2}\left(\frac{\alpha}{\theta} - \frac{d\alpha}{d\theta}\right) = \langle\kappa\rangle - \kappa. \tag{B.8}$$

The eigenvectors of M^{-1} point in the radial and tangential directions, with a radial eigenvalue of $\lambda_+ = 1-\kappa+\gamma = 1-d\alpha/d\theta$ and a tangential eigenvalue of $\lambda_- = 1-\kappa-\gamma = 1-\alpha/\theta = 1-\langle\kappa\rangle$. If either one of these eigenvalues is zero, the magnification diverges and we are on either the radial or tangential critical curve. If we can resolve the images, we will see the images radially magnified near the radial critical curve and tangentially magnified near the tangential critical curve. For example, all the quasar host galaxies seen in Figs. B.3 and B.4 lie close to the tangential critical line and are stretched tangentially to form partial or complete Einstein rings. The signs of the eigenvalues $\lambda_\pm$ give the parities of the images and the type of time delay extremum associated with the images. If both eigenvalues are positive, the image is a minimum. If both are negative, the image is a maximum. If one is positive and the other negative, the image is a saddle point. The inverse of the total magnification $\mu^{-1} = |M^{-1}|$ is the product of the eigenvectors, so it is positive for minima and maxima and negative for saddle points. The signs of the eigenvalues are referred to as the partial parities of the images, while the sign of the total magnification is referred to as the total parity.

It is useful to use simple examples to illustrate the behavior of circular lenses for different density profiles. In most previous lensing reviews, the examples are based on lenses with finite core radii. However, most currently popular models of galaxies and clusters have central density cusps rather than core radii, so we will depart from historical practice and focus on the *power-law lens* (e.g. Evans & Wilkinson 1998). Suppose, in three dimensions, that the lens has a density distribution $\rho \propto r^{-n}$. Such a lens will produce deflections of

$$\alpha(\theta) = b\left(\frac{\theta}{b}\right)^{2-n} \tag{B.9}$$

as shown in Fig. B.9, with convergence and shear profiles

$$\kappa(\theta) = \frac{3-n}{2}\left(\frac{\theta}{b}\right)^{1-n} \quad \text{and} \quad \gamma(\theta) = \frac{n-1}{2}\left(\frac{\theta}{b}\right)^{1-n}. \tag{B.10}$$

The power law lenses cover most of the simple, physically interesting models. The *point mass* lens is the limit $n \to 3$, with deflection $\alpha = b^2/\theta$, convergence $\kappa = 0$ (with a central singularity) and shear $\gamma = b^2/r^2$. The *singular isothermal sphere (SIS)* is the case with $n = 2$. It has a constant deflection $\alpha = b$, and equal convergence and shear $\kappa = \gamma = b/2\theta$. A *uniform critical sheet* is the limit $n \to 1$ with $\alpha = \theta$, $\kappa = 1$ and $\gamma = 0$. Models with $n \to 3/2$ have the cusp exponent of the Moore (1998) halo model. The popular $\rho \propto 1/r$ NFW (Navarro, Frenk & White 1996, see §B.4.1) density cusps are not quite the same as the $n \to 1$ case because the projected surface density of a $\rho \propto 1/r$ cusp has $\kappa \propto \ln\theta$ rather than a constant. Nonetheless, the behavior of the power law models as $n \to 1$ will be very similar to the NFW model if the lens is dominated by the central cusp. The central regions of galaxies probably act like cusps with $1 \lesssim n \lesssim 2$.

The tangential magnification eigenvalue of these models is

$$\lambda_- = 1 - \kappa - \gamma = 1 - \frac{\alpha}{\theta} = 1 - \langle\kappa\rangle = 1 - (\theta/b)^{1-n} \tag{B.11}$$

which is always equal to zero at $\theta = b \equiv \theta_E$. This circle defines the tangential critical curve or Einstein (ring) radius of the lens. We normalized the models in this fashion because the Einstein radius is usually the best-determined parameter of any lens model, in the sense that all successful models will find nearly the same Einstein radius (e.g. Kochanek 1991a, Wambsganss & Paczynski 1994). The source position corresponding to the tangential critical curve is the origin ($\beta = 0$), and the reason the magnification diverges is that a point source at the origin is converted into a ring on the tangential critical curve leading to a divergent ratio between the "areas" of the source and the image. The other important point to notice is that the mean surface density inside the tangential critical radius is $\langle\kappa\rangle \equiv 1$ independent of the model. This is true of any circular lens. With the addition of angular structure it is not strictly true, but it is a very good approximation unless the mass distribution is very flattened. The definition of b in terms of the properties of the lens galaxy will depend on the particular profile. For example, in a point mass lens ($n \to 3$), $b^2 = (4GM/c^2 D_d^{ang})(D_{ds}/D_s)$ where M is the mass, while in an SIS lens ($n = 2$), $b = 4\pi(\sigma_v/c)^2 D_{ds}/D_s$ where σ_v is the (1D) velocity dispersion of the lens. For the other profiles, b can be defined in terms of some velocity dispersion or mass estimate for the lens, as we will discuss later in §B.4.9 and §B.6. The radial magnification eigenvalue of these models is

$$\lambda_+ = 1 - \kappa + \gamma = 1 - \frac{d\alpha}{d\theta} = 1 - (2-n)(\theta/b)^{1-n} \tag{B.12}$$

which can be zero only if $n < 2$. If $n < 2$ the deflection goes to zero at the origin and the lens has a radial critical curve at $\theta = b(2-n)^{1/(n-1)} < b$

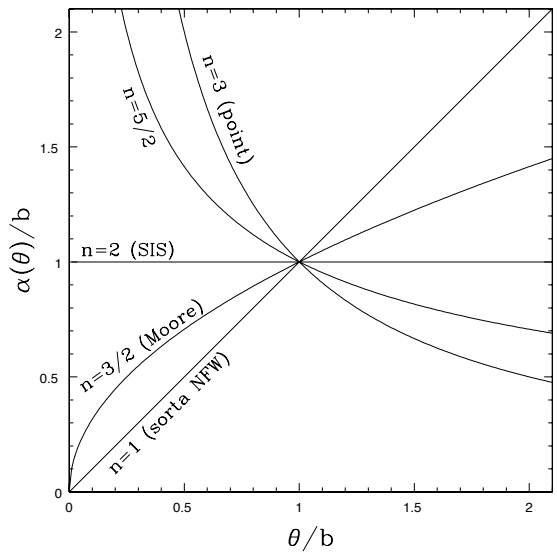


Fig. B.9. The bending angles of the power law lens models. Profiles more centrally concentrated ($n > 2$) than the SIS ($n = 2$), have divergent central deflections, while profiles more extended ($n < 2$) than SIS have deflection profiles that become zero at the center of the lens. The $n = 1$ model is not quite an NFW model because the surface density is constant rather than logarithmic.

interior to the tangential critical curve. Models with $n \geq 2$ have constant ($n = 2$) or rising deflection profiles as we approach the lens center and have negative derivatives $d\alpha/d\theta$ at all radii.

A nice property of circular lenses is that they allow simple graphical solutions of the lens equation for arbitrary deflection profiles. There are two parts to the graphical solution – the first is to determine the radial positions θ_i of the images given a source position β, and the second is to determine the magnification by comparing the area of the images to the area of the source. Recall first, that by symmetry, all the images must lie on a line passing through the source and the lens. Let θ now be a signed radius that is positive along this line on one side of the lens and negative on the other. The lens equation (Eqn. B.4) along the line is simply

$$\frac{\theta}{|\theta|}\alpha(|\theta|) = \theta - \beta \tag{B.13}$$

where we have rearranged the terms to put the deflection on one side and the image and source positions on the other. One side of the equation is the bend angle (Fig. B.9), while the other side of the equation, $\theta - \beta$, is simply a

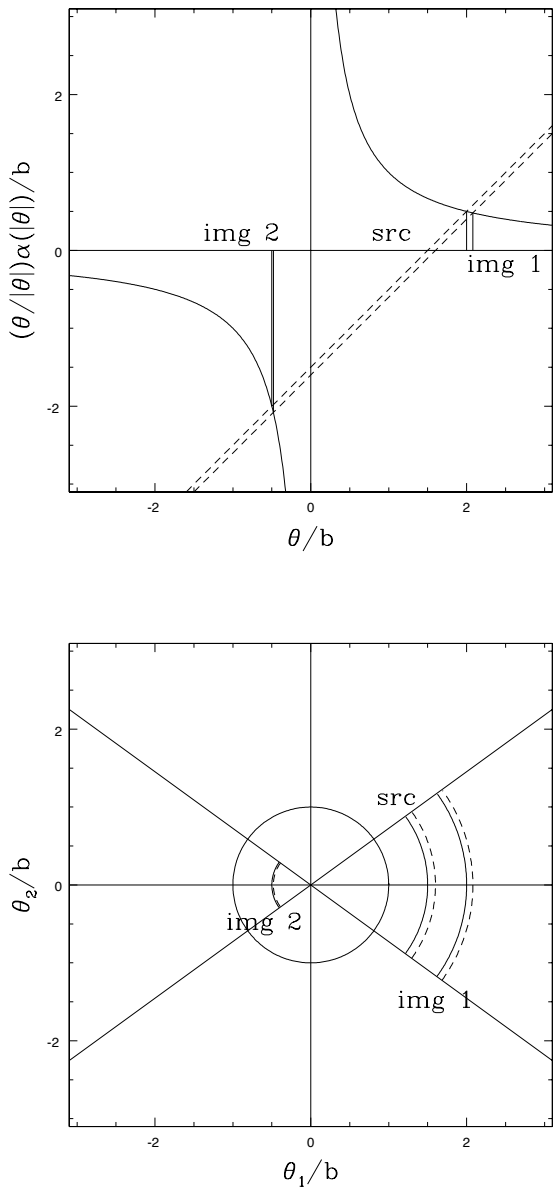


Fig. B.10. Graphical solutions for the point mass ($n = 3$) lens. The top panel shows the graphical solution for the radial positions of the images, and the bottom panel shows the graphical solution for the image structure. Note the strong radial demagnification of image 2 produced by the falling deflection profile.

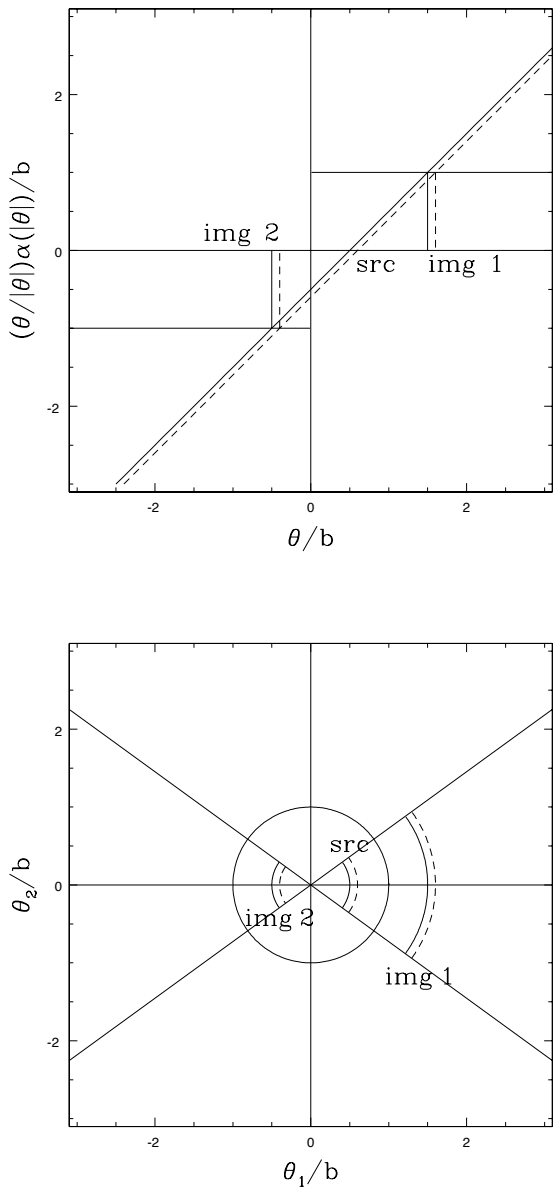


Fig. B.11. Graphical solutions for the SIS ($n = 2$) lens when $\beta < b$ and there are two images.

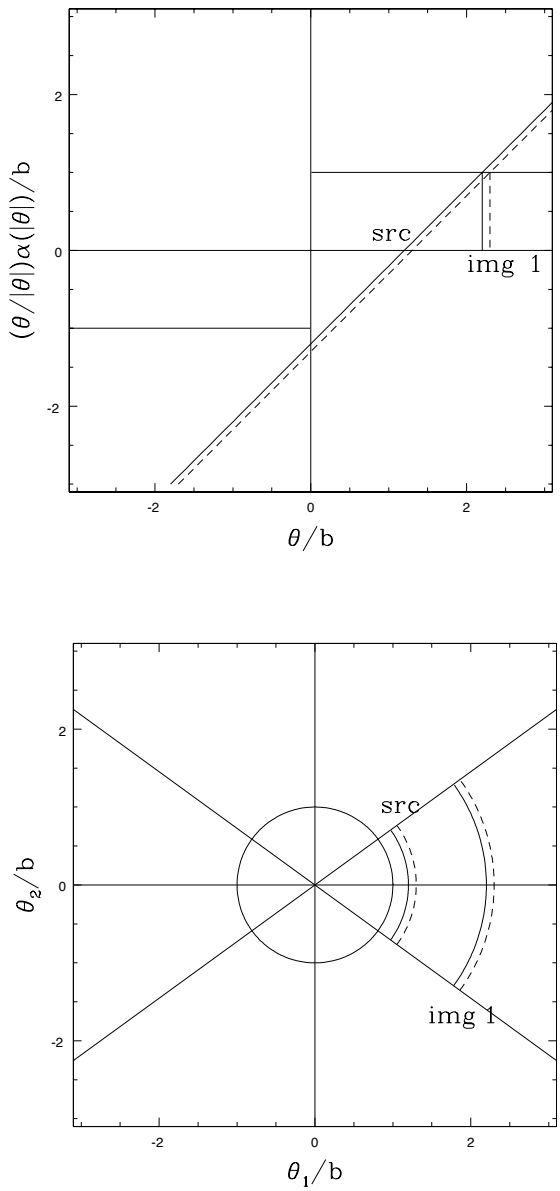


Fig. B.12. Graphical solutions for the SIS ($n = 2$) lens when $\beta > b$ and there is only one image.

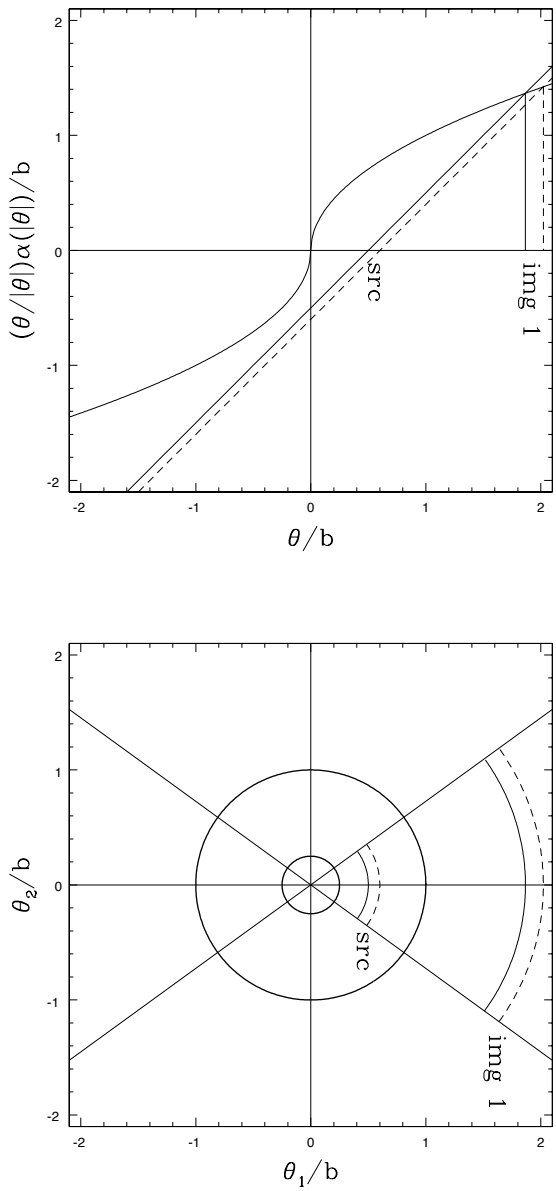


Fig. B.13. Graphical solutions for the Moore profile cusp ($n = 3/2$) lens when $\beta > b/4$ and there is only one image.

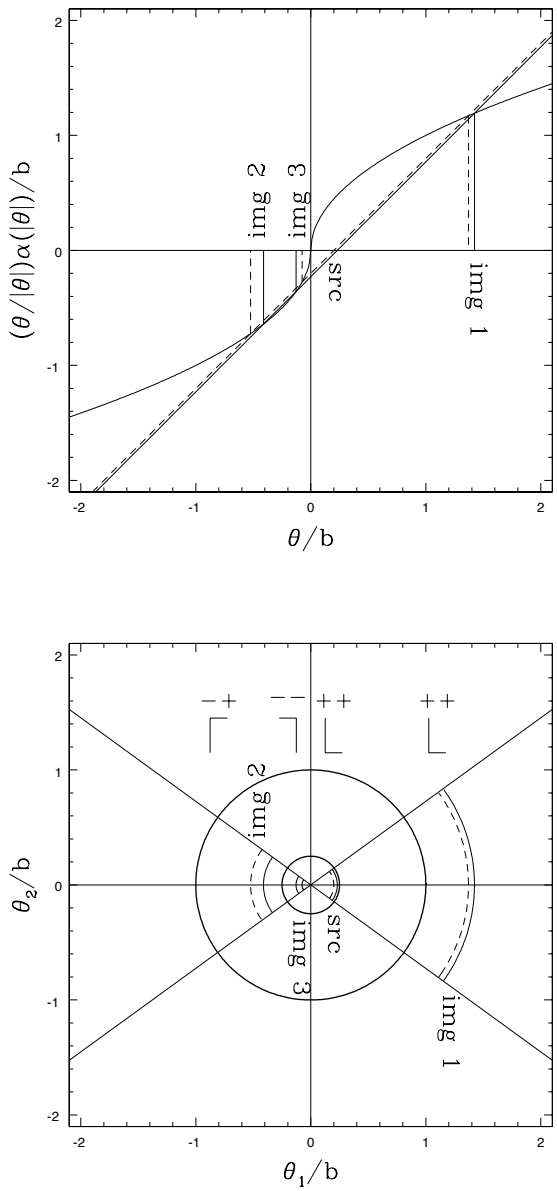


Fig. B.14. Graphical solutions for the Moore profile cusp ($n = 3/2$) lens when $\beta < b/4$ and there are three images. At the top of the lower panel we illustrate the geometric meaning of the image partial parities defined by the signs of the magnification tensor eigenvalues (see text).

line of unit slope passing through the source position β. The solutions to the lens equation for any source position β are the radii θ_i where the line crosses the curve.

For understanding any observed lens, it is always useful to first sketch where the critical lines must lie. Recall from the discussion of caustics in Part 1, that images are always created and destroyed on critical lines as the source crosses a caustic, so the critical lines and caustics define the general structure of the lens. All our power law models have a tangential critical line at $\theta = b$, which is the solution $\alpha(b) = b$ and corresponds to the source position $\beta = 0$. The origin, as the projection of the critical curve onto the source plane, is the tangential caustic (strictly speaking a degenerate pseudo-caustic) corresponding to the critical line. A point source at the origin is transformed into an Einstein ring of radius $\theta_E = b$.

The second step of the graphical construction is to determine the angular structure of the image. For simplicity, suppose the source is an arc with radial width $\Delta\beta$ and angular width $\Delta\chi$. By symmetry, the angle subtended by an image relative to the lens center must be the same as that subtended by the source. For an image at θ_i and a source at β, the tangential extent of the image is $|\theta_i|\Delta\chi$ while that of the source is $\beta\Delta\chi$. The tangential magnification of the image is simply $|\theta_i|/\beta = (1 - |\alpha(\theta_i)/\theta_i|)^{-1}$ after making use of the lens equation (Eqn. B.13), and this is identical to the tangential magnification eigenvalue (Eqn. B.11). The thickness of the arc requires finding the image radii for the inner and outer edges of the source, $\theta_i(\beta)$ and $\theta_i(\beta + \Delta\beta)$. The ratio of the thickness of the two arcs is the radial magnification,

$$\frac{\theta_i(\beta + \Delta\beta) - \theta_i(\beta)}{\Delta\beta} \simeq \frac{d\theta}{d\beta} = \left(1 - \frac{d\alpha}{d\theta}\right)^{-1}, \tag{B.14}$$

and this is simply the inverse of the radial eigenvalue of the magnification matrix (Eqn. B.12) where we have taken the derivative of the lens equation (Eqn. B.13) with respect to the source position to obtain the final result. Thus, the tangential magnification simply reflects the fact that the angle subtended by the source is the angle subtended by the image, while the radial magnification depends on the slope of the deflection profile with declining deflection profiles ($d\alpha/d\theta < 0$) demagnifying the source and rising profiles magnifying the source.

In Fig. B.10 we illustrate this for the point mass lens ($n \to 3$). From the shape of the deflection profile, it is immediately obvious that there will be only two images, one on each side of the lens. If we assume $\beta > 0$, the first image is a minimum located at

$$\theta_1 = \frac{1}{2}\left(\beta + \sqrt{\beta^2 + 4b^2}\right) \tag{B.15}$$

with $\theta_1 > \theta_E$ and positive magnification

$$\mu_1 = \frac{1}{4}\left(\frac{\beta}{\sqrt{\beta^2+4b^2}} + \frac{\sqrt{\beta^2+4b^2}}{\beta} + 2\right) > 0, \tag{B.16}$$

while the second image is a saddle point located at

$$\theta_2 = \frac{1}{2}\left(\beta - \sqrt{\beta^2+4b^2}\right) \tag{B.17}$$

with $-\theta_E < \theta_2 < 0$ and negative magnification

$$\mu_2 = -\frac{1}{4}\left(\frac{\beta}{\sqrt{\beta^2+4b^2}} + \frac{\sqrt{\beta^2+4b^2}}{\beta} - 2\right) < 0. \tag{B.18}$$

As the source approaches the tangential caustic ($\beta \to 0$) the magnifications of both images diverge as β^{-1} and the image radii converge to θ_E. As the source moves to infinity, the magnification of the first image approaches unity and its position approaches that of the source, while the second image is demagnified by the factor $(1/2)(b/\beta)$ and converges to the position of the lens. The image separation

$$\Delta\theta = |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| = 2b\sqrt{1+\beta^2/4b^2} \geq 2b \tag{B.19}$$

is always larger than the diameter of the Einstein ring and the total magnification

$$|\mu_1| + |\mu_2| = \frac{2b^2+\beta^2}{\beta\sqrt{\beta^2+4b^2}} \geq 1 \tag{B.20}$$

is the characteristic light curve expected for isolated Galactic microlensing events (see Part 4). The point mass lens has one peculiarity that makes it different from extended density distributions like galaxies in that it has two images independent of the impact parameter of the source and no radial caustic. This is a characteristic of any density distribution with a divergent central deflection ($n > 2$).

The SIS ($n = 2$) model is the "standard" lens model for galaxies. Figs. B.11 and B.12 show the geometric constructions for the images of an SIS lens. If $0 < \beta < b$, then the SIS lens also produces two images (Fig. B.11). The first image is a minimum located at

$$\theta_1 = \beta + b \quad \text{with } \theta_1 > b \text{ and positive magnification} \quad \mu_1 = 1 + b/\beta \tag{B.21}$$

and the second image is a saddle point located at

$$\theta_2 = \beta - b \quad \text{with } -b < \theta_2 < 0 \text{ and negative magnification} \quad \mu_2 = 1 - b/\beta. \tag{B.22}$$

The image separation $|\theta_1 - \theta_2| = 2b$ is constant, and the total magnification $|\mu_1| + |\mu_2| = 2b/\beta$ is a simple power law. The magnification produced by

an SIS lens is purely tangential since the radial magnification is unity. If, however, $\beta > b$, then there is only one image, corresponding to the minimum located on the same side of the lens as the source (see Fig. B.12). This boundary on the source plane at $\beta = b$ between having two images at smaller radii and only one image at larger radii is a radial (pseudo)-caustic that can be thought of as being associated with a radial critical curve at the origin. It is a pseudo-caustic because there are neither images nor a divergent magnification associated with it.

Historically the next step is to introduce a core radius to have a model with a true radial critical line and caustic (see Part 1, Blandford & Kochanek 1987b, Kochanek & Blandford 1987, Kovner 1987a, Hinshaw & Krauss 1987, Krauss & White 1992, Wallington & Narayan 1993, Kochanek 1996a). Instead we will consider the still softer power law model with $n = 3/2$, which would correspond to the central exponent of the "Moore" profile proposed for CDM halos (Moore et al. 1998). As Fig. B.13 shows, there is only one solution for $|\beta| > b/4$, a minimum located at

$$\theta_1 = \frac{1}{2}\left(b + 2\beta + \sqrt{b + 4\beta}\right) \tag{B.23}$$

and with $\theta_1 > b$ assuming β is positive. The magnification expressions are too complex to be of much use, but the magnification μ_1 diverges at $\theta = b$ when the source is on the tangential pseudo-caustic at $\beta = 0$. As Fig. B.14 shows, we find two additional images once $|\beta| < b/4$. The first additional image is a saddle point located at

$$\theta_2 = \frac{1}{2}\left(-b + 2\beta - \sqrt{b + 4\beta}\right) \tag{B.24}$$

with $-b < \theta_2 < -b/4$, which has a negative magnification that diverges at both $\theta_2 = -b$ (the tangential critical curve) and $\theta_2 = -b/4$. This latter radius defines the radial critical curve where the magnification diverges because the radial magnification eigenvalue $1 - \kappa + \gamma = 1 - d\alpha/d\theta = 0$ at radius $\theta = b/4$. The third image is a maximum located at

$$\theta_3 = \frac{1}{2}\left(-b + 2\beta + \sqrt{b + 4\beta}\right) \tag{B.25}$$

with $-b/4 < \theta_2 < 0$ and a positive magnification that diverges on the radial critical curve. As we move the source outward from the center we would see images 2 and 3 approach each other, merging on the radial critical line where they would have divergent magnifications, and then vanishing to leave only image 1. We would see the same pattern if instead of softening the exponent we had followed the traditional path and added a core radius to the SIS model. With a finite core radius the central deflection profile would pass through zero, and this would introduce a radial critical curve and a third image which would be a maximum of the time delay surface.

In Fig. B.14 we also illustrate the geometric meaning of the partial parities (the signs of the magnification eigenvalues). A source structure (the L) defines the reference shape. Image 1 is a minimum with positive partial parities (++) defined by the signs of the tangential and radial eigenvalues. The orientation of image 1 is the same as the source. Image 2 is a saddle point with mixed partial parities $(-+)$ because the tangential eigenvalue is negative while the radial eigenvalue is positive. This means that the image is inverted in the tangential direction relative to the source. Image 3 is a maximum with negative partial parties $(--)$, so the image is inverted in both the radial and tangential directions relative to the source. The total parity, the product of the partial parities, is positive for maxima and minima so the orientation of the image can be produced by rotating the source. The total parity of the saddle point image is negative, so its orientation cannot be produced by a rotation of the source.

B.3.3 Non-Circular Lenses

The tangential pseudo-caustic at the origin producing Einstein ring images is unstable to the introduction of any angular structure into the gravitational potential of the lens. There are two generic sources of angular perturbations. The first source of angular perturbations is the ellipticity of the lens galaxy. What counts here is the ellipticity of the gravitational potential rather than of the surface density. For a lens with axis ratio q, ellipticity $\epsilon = 1 - q$, or eccentricity $e = (1 - q^2)^{1/2}$, the ellipticity of the potential is usually $\epsilon_\Psi \sim \epsilon/3$ – potentials are always rounder than densities. The second source of angular perturbations is tidal perturbations from any nearby objects. This is frequently called the "external shear" or the "tidal shear" because it can be modeled as a linear shearing of the deflections. In all known lenses, quadrupole perturbations (i.e. $\Psi \propto \cos(2\chi)$ where χ is the azimuthal angle) dominate – higher order multipoles are certainly present and they can be quantitatively important, but they are smaller. For example, in an ellipsoid the amplitude of the $\cos 2m\chi$ multipole scales as ϵ_Ψ^m (see §B.4.4 and §B.8).

Unfortunately, there is no example of a non-circular lens that can be solved in full generality unless you view the nominally analytic solutions to quartic equations as helpful. We can make the greatest progress for the case of an SIS in an external (tidal) shear field. Tidal shear is due to perturbations from nearby objects and its amplitude can be determined by Taylor expanding its potential near the lens (see Part 1 and §B.4). Consider a lens with Einstein radius θ_E perturbed by an object with effective lens potential Ψ a distance θ_p away. For $\theta_E \ll \theta_p$ we can Taylor expand the potential of the nearby object about the center of the primary lens, dropping the leading two terms.[3] This

[3] The first term, a constant, gives an equal contribution to the time delays of all the images, so it is unobservable when all we can measure is relative delays. The second term is a constant deflection, which is unobservable when all we can measure is relative deflections.

leaves, as the first term with observable consequences,

$$\Psi(\theta) \simeq \frac{1}{2}\boldsymbol{\theta}\cdot\boldsymbol{\nabla}\boldsymbol{\nabla}\Psi\cdot\boldsymbol{\theta} = \frac{1}{2}\kappa_p\theta^2 - \frac{1}{2}\gamma_p\theta^2\cos 2(\chi-\chi_p) \tag{B.26}$$

where κ_p is the surface density of the perturber at the center of the lens galaxy and $\gamma_p > 0$ is the tidal shear from the perturber. If the perturber is an SIS with critical radius b_p and distance θ_p from the primary lens, then $\kappa_p = \gamma_p = b_p/2\theta_p$. With this normalization, the angle χ_p points toward the perturber. For a circular lens, the shear $\gamma_p = \langle\kappa\rangle - \kappa$ can be expressed in terms of the surface density of the perturber, and it is larger (smaller) than the convergence if the density profile is steeper (shallower) than isothermal.

The effects of κ_p are observable only if we measure a time delay or have an independent estimate of the mass of the lens galaxy, while the effects of the shear are easily detected from the relative positions of the lensed images (see Part 1). Consider, for example, one component of the lens equation including an extra convergence,

$$\beta_1 = \theta_1(1-\kappa_p) - d\Psi/d\theta_1 \tag{B.27}$$

and then simply divide by $1-\kappa_p$ to get

$$\beta_1/(1-\kappa_p) = \theta_1 - (d\Psi/d\theta_1)/(1-\kappa_p). \tag{B.28}$$

The rescaling of the source position $\beta_1/(1-\kappa_p)$ has no consequences since the source position is not an observable quantity, while the rescaling of the deflection is simply a change in the mass of the lens. This is known as the "mass sheet degeneracy" because it corresponds to adding a constant surface density sheet to the lens model (Falco, Gorenstein & Shapiro 1985, see Part 1), and it is an important systematic problem for both strong lenses and cluster lenses (see Part 3).

Thus, while the extra convergence can be important for the quantitative understanding of time delays or lens galaxy masses, it is only the shear that introduces qualitatively new behavior to the lens equations. The effective potential of an SIS lens in an external shear is $\Psi = b\theta + (\gamma/2)\theta^2\cos 2\chi$ leading to the lens equations

$$\begin{aligned}\beta_1 &= \theta_1(1-\gamma) - b\theta_1/|\boldsymbol{\theta}|\\ \beta_2 &= \theta_2(1+\gamma) - b\theta_2/|\boldsymbol{\theta}|\end{aligned} \tag{B.29}$$

where for $\gamma > 0$ the perturber is due North (or South) of the lens. The inverse magnification is

$$\mu^{-1} = 1-\gamma^2 - \frac{b}{\theta}(1-\gamma\cos 2\chi) \tag{B.30}$$

where $\boldsymbol{\theta} = (\theta_1,\theta_2) = \theta(\cos\chi,\sin\chi)$.

The first step in any general analysis of a new lens potential is to locate the critical lines and caustics. In this case we can easily solve $\mu^{-1} = 0$ to find

that the tangential critical line

$$\theta = b\frac{1-\gamma\cos 2\chi}{1-\gamma^2} \tag{B.31}$$

is an ellipse whose axis ratio is determined by the amplitude of the shear γ and whose major axis points toward the perturber. We call it the tangential critical line because the associated magnifications are nearly tangential to the direction to the lens galaxy and because it is a perturbation to the Einstein ring of a circular lens. The tangential caustic, the image of the critical line on the source plane, is a curve called an astroid (Fig.B.15, it is not a "diamond" despite repeated use of the term in the literature). The parametric expression for the astroid curve is

$$\beta_1 = -\frac{2b\gamma}{1+\gamma}\cos^3\chi = -\beta_+\cos^3\chi \qquad \beta_2 = +\frac{2b\gamma}{1-\gamma}\sin^3\chi = \beta_-\sin^3\chi \tag{B.32}$$

where the parameter χ is the same as the angle appearing in the critical curve (Eqn. B.31) and we have defined $\beta_\pm = 2b\gamma/(1\pm\gamma)$ for the locations of the cusp tips on the axes. The astroid consists of 4 cusp caustics on the symmetry axes of the lens connected by fold caustics with a major axis pointing toward the perturber. Like the SIS model without any shear, the origin plays the role of the radial critical line and there is a circular radial pseudo-caustic at $\beta = b$.

As mentioned earlier, there is no useful general solution for the image positions and magnifications. We can, however, solve the equations for a source on one of the symmetry axes of the lens. For example, consider a solution on the minor axis of the lens ($\beta_2 = 0$ for $\gamma > 0$). There are two ways of solving the lens equation to satisfy the criterion. One is to put the images on the same axis ($\theta_2 = 0$) and the other is to place them on the arc defined by $0 = 1+\gamma-b/\theta$. The images with $\theta_2 = 0$ are simply the SIS solutions corrected for the effects of the shear. Image 1 is defined by

$$\theta_1 = \frac{\beta_1+b}{1-\gamma} \quad \text{with} \quad \mu^{-1} = \left(1-\gamma^2\right)\frac{\beta_+ + \beta_1}{b+\beta_1} \tag{B.33}$$

and image 2 is defined by

$$\theta_1 = \frac{\beta_1-b}{1-\gamma} \quad \text{with} \quad \mu^{-1} = \left(1-\gamma^2\right)\frac{\beta_+ - \beta_1}{b-\beta_1} \tag{B.34}$$

Image 1 exists if $\beta_1 > -b$, it is a saddle point for $-b < \beta_1 < -\beta_+$ and it is a minimum for $\beta_1 > -\beta_+$. Image 2 has the reverse ordering. It exists for $\beta_1 < b$, it is a saddle point for $\beta_+ < \beta_1 < b$ and it is a minimum for $\beta_1 < \beta_+$. The magnifications of both images diverge when they are on the tangential critical line ($\beta_1 = -\beta_+$ for image 1 and $\beta_1 = +\beta_+$ for image 2) and approach zero as they move into the core of the lens ($\beta_1 \to -b$ for image 1 and $\beta_1 \to +b$

for image 2). These two images shift roles as the source moves through the origin. The other two solutions are both saddle points, and they exist only if the source lies inside the astroid ($|\beta_1| < \beta_+$ along the axis). The positions of images 3 (+) and 4 (−) are

$$\theta_1 = -\frac{\beta_1}{2\gamma} \qquad \theta_2 = \pm\frac{b}{1+\gamma}\left[1-\left(\frac{\beta_1}{\beta_+}\right)^2\right]^{1/2} \tag{B.35}$$

and they have equal magnifications

$$\mu^{-1} = -2\gamma(1+\gamma)\left[1-\left(\frac{\beta_1}{\beta_+}\right)^2\right]. \tag{B.36}$$

The magnifications of the images diverge when the source reaches the cusp tip ($|\beta_1| = \beta_+$) and the image lies on the tangential critical curve.

Thus, if we start with a source at the origin we can follow the changes in the image structure (see Fig. B.15, B.16). With the source at the origin we see 4 images on the symmetry axes with reasonably high magnifications, $\sum|\mu_i| = (2/\gamma)/(1-\gamma^2) \sim 10$. It is a generic result that the least magnified four-image system is found for an on-axis source, and this configuration has a total magnification of order the inverse of the ellipticity of the gravitational potential. As we move the source toward the tip of the cusp ($\beta \to \beta_+$, Fig. B.15), image 1 simply moves out along the symmetry axis with slowly dropping magnification, while images 2, 3 and 4 move toward a merger on the tangential critical curve at $\boldsymbol{\theta} = (-\beta_+, 0)$. Their magnifications steadily rise and then diverge when the source reaches the cusp. If we move the source further outward we find only images 1 and 2 with 1 moving outward and 2 moving inward toward the origin. As it approaches the origin, image 2 becomes demagnified and vanishes when $\beta \to b$. Had we done the same calculation on the major axis (Fig. B.16), there is a qualitative difference. As we moved image 1 outward along the β_2 axis, image 3 and 4 would merge with image 1 when the source reaches the tip of the cusp at $\beta_2 = \beta_-$ rather than with image 2.

Unfortunately once we move the source off a symmetry axis, there is no simple solution. It is possible to find the locations of the remaining images given that two images have merged on the critical line, and this is useful for determining the mean magnifications of the lensed images, a point we will return to when we discuss lens statistics in §B.6. Here we simply illustrate (Fig. B.17) the behavior of the images when we move the source radially outward from the origin away from the symmetry axes. Rather than three images merging on the tangential critical line as the source approaches the tip of a cusp, we see two images merging as the source approaches the fold caustic of the astroid. This difference, two images merging versus three images merging, is a generic difference between folds and cusps as discussed in Part 1. All images in these four-image configurations are restricted to an annulus of

width $\sim \gamma b$ around the critical line, so the mean magnification of all four image configurations is also of order γ^{-1} (see Finch et al. 2002).

There is one more possibility for the caustic structure of the lens if the external shear is large enough. For $1/3 < |\gamma| < 1$, the tip of the astroid caustic extends outside the radial caustic, as shown if Fig. B.18. This allows a new image geometry, known as the cusp or disk geometry, where we see three images straddling the major axis of a very flattened potential. It is associated with the caustic region inside the astroid caustic but outside the radial caustic. This configuration appears to be rare for lenses produced by galaxies, with APM08279+5255 as the only likely candidate, but relatively more common in clusters. The difference is that clusters tend to have shallower density profiles than galaxies, which shrinks the radial caustics relative to the tangential caustics to allow more cross section for this image configuration and lower ellipticity thresholds before it becomes possible (Oguri & Keeton 2004 most recently, but also see Kochanek & Blandford 1987, Kovner 1987a, Wallington & Narayan 1993).

In general, it is far more difficult to analyze ellipsoidal lenses, in part because few ellipsoidal lenses have analytic expressions for their deflections. The exception is the isothermal ellipsoid (Kassiola & Kovner 1993, Kormann, Schneider & Bartelmann 1994, Keeton & Kochanek 1998), including a core radius s, which is both analytically tractable and generally viewed as the most likely average mass distribution for gravitational lenses. The surface density of the isothermal ellipsoid

$$\kappa = \frac{1}{2}\frac{b}{\omega} \quad \text{where} \quad \omega^2 = q^2(\theta_1^2 + s^2) + \theta_2^2 \tag{B.37}$$

depends on the axis ratio q and the core radius s. For $q = 1 - \epsilon < 1$ the major axis is the θ_1 axis and s is the major axis core radius. The deflections produced by this lens are remarkably simple,

$$\alpha_1 = \frac{b}{\sqrt{1-q^2}}\tan^{-1}\left[\frac{\theta_1\sqrt{1-q^2}}{\omega+s}\right] \quad \text{and} \quad \alpha_2 = \frac{b}{\sqrt{1-q^2}}\tanh^{-1}\left[\frac{\theta_2\sqrt{1-q^2}}{\omega+q^2 s}\right]. \tag{B.38}$$

The effective lens potential is cumbersome but analytic,

$$\Psi = \boldsymbol{\theta}\cdot\boldsymbol{\alpha} - bs\ln\left[(\omega+s)^2 + (1-q^2)\theta_1^2\right]^{1/2}, \tag{B.39}$$

the magnification is simple

$$\mu^{-1} = 1 - \frac{b}{\omega} - \frac{b^2 s}{\omega\left[(\omega+s)^2 + (1-q^2)\theta_1^2\right]} \tag{B.40}$$

and becomes even simpler in the limit of a singular isothermal ellipsoid (SIE) with $s = 0$ where $\mu^{-1} \to 1 - b/\omega$. In this case, contours of surface density κ are also contours of the magnification, and the tangential critical line is the $\kappa =$

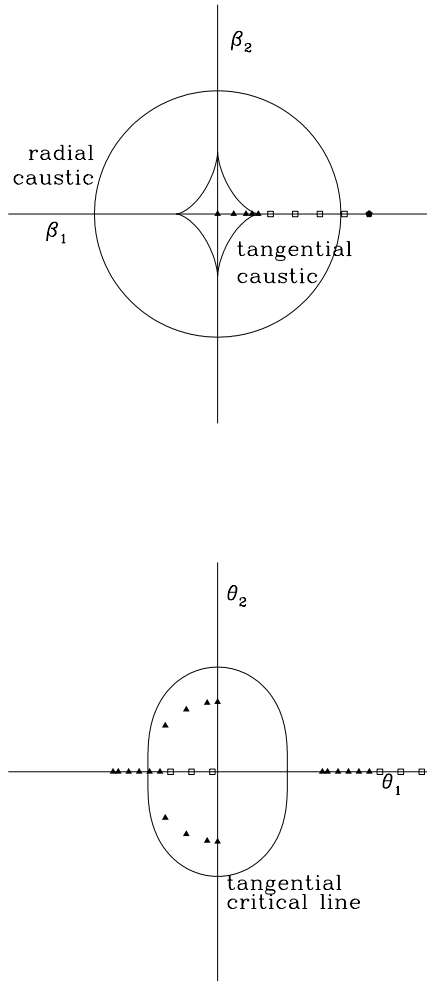


Fig. B.15. Example of a minor axis cusp on the source (top) and image (bottom) planes. When the source is inside both the radial and tangential caustics (triangles) there are four images. As the source moves toward the cusp, three of the images head towards a merger on the critical line and become highly magnified to leave only one image once the source crosses the cusp and lies between the two caustics (open squares). In a minor axis cusp, the image surviving the cusp merger is a saddle point interior to the critical line. As the source approaches the radial caustic, one image approaches the center of the lens and then vanishes as the it crosses the caustic to leave only one image (pentagons).

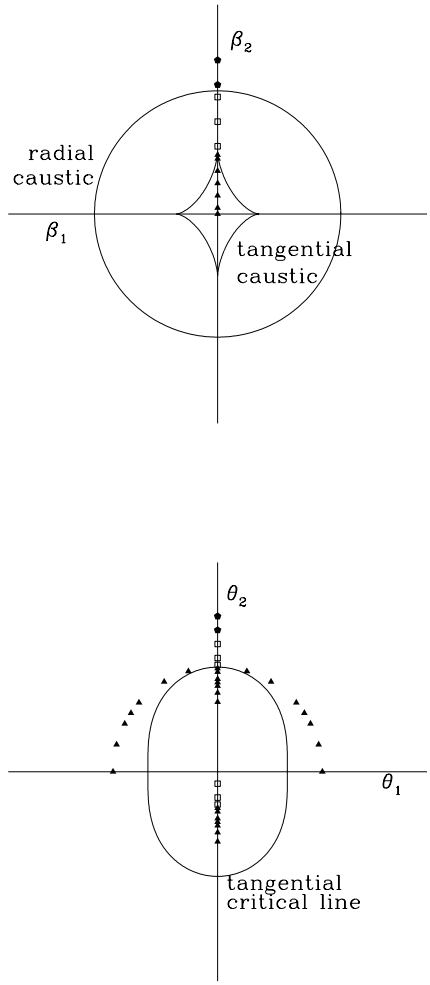


Fig. B.16. Example of a major axis cusp on the source (top) and image (bottom) planes. When the source is inside both the radial and tangential caustics (triangles) there are four images. As the source moves toward the cusp, three of the images head towards a merger on the critical line and become highly magnified to leave only one image once the source crosses the cusp and lies between the two caustics (open squares). In a major axis cusp, the image surviving the cusp merger is the minimum corresponding to the image we would see in the absence of a lens. As the source approaches the radial caustic, one image approaches the center of the lens and then vanishes as the source crosses the caustic to leave only one image (pentagons).

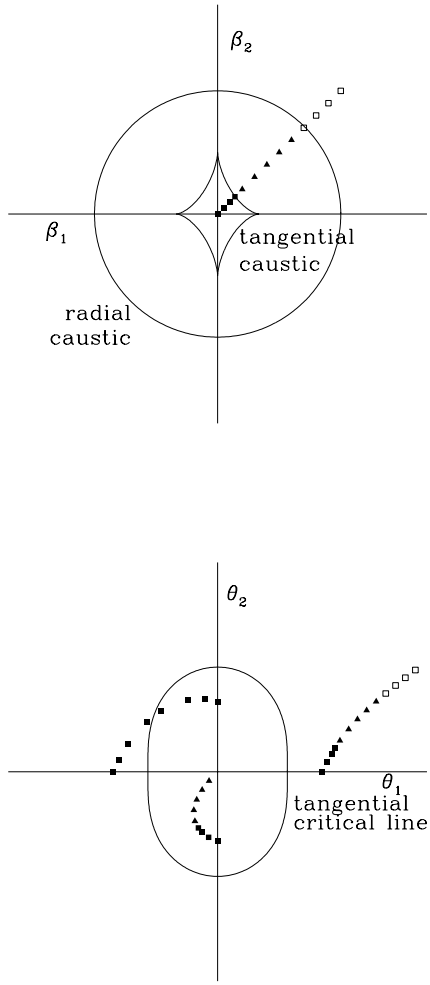


Fig. B.17. Example of a fold merger on the source (top) and image (bottom) planes. When the source is inside both the radial and tangential caustics (filled squares) there are four images. As the source crosses the tangential caustic, two images merge, become highly magnified and then vanish, leaving only two images (triangles) when the source is outside the tangential caustic but inside the radial caustic. As the source approaches the radial caustic, one image moves into the center of the lens and then vanishes when the source crosses the radial caustic to leave only one image when the source is outside both caustics (open squares).

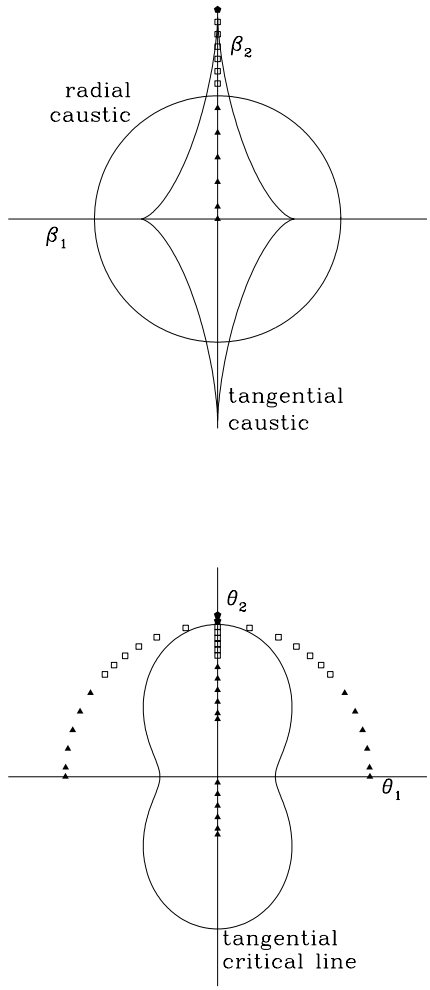


Fig. B.18. Example of a cusp or disk image geometry on the source (top) and image (bottom) planes. The shear is high enough to make the tangential caustic extend outside the radial caustic. For a source inside both caustics (triangles) we see a standard four-image geometry as in Fig. B.16. However, for a source outside the radial caustic but inside the tangential caustic (squares) we have three images all on one side of the lens. This is known as the cusp geometry because it is always associated with cusps, and the disk geometry because flattened disks are the only natural way to produce them. Once the source is outside the cusp tip (pentagon), a single image remains.

1/2 isodensity contour just as for the SIS model. The critical radius scale b can be related to the circular velocity in the plane of the galaxy relatively easily. For an isothermal sphere we have that $b_{SIS} = 4\pi(\sigma_v/c)^2 D_{ds}/D_s$ where the circular velocity is $v_c = \sqrt{2}\sigma_v$. For the projection of a three-dimensional (3D) oblate ellipsoid of axis ratio q_3 and inclination i, so that $q^2 = q_3^2 \cos^2 i + \sin^2 i$, the deflection scale is $b = b_{SIS}(e_3/\sin^{-1} e_3)$ where $e_3 = \sqrt{1-q_3^2}$ is the eccentricity of 3D mass distribution. In the limit that $q_3 \to 0$ the model becomes a Mestel (1963) disk, the infinitely thin disk producing a flat rotation curve, and $b = 2b_{SIS}/\pi$ (see §B.4.9 and Keeton, Kochanek & Seljak 1997, Keeton & Kochanek 1998, Chae 2003). At least for the case of a face-on disk, at fixed circular velocity you get a smaller Einstein radius as you make the 3D distribution flatter because a thin disk requires less mass to produce the same circular velocity.

We can generate several other useful models from the isothermal ellipsoids. For example, steeper ellipsoidal density distributions can be derived by differentiating with respect to s^2. The most useful of these is the first derivative with $\kappa \propto \omega^{-3/2}$ which is related to the Kuzmin (1956) disk (see Kassiola & Kovner 1993, Keeton & Kochanek 1998). It is also easy to generate models with flat inner rotation curves and truncated halos by taking the difference of two isothermal ellipsoids. In particular if $\kappa(s)$ is an isothermal ellipsoid with core radius s, the model

$$\kappa = \kappa(s) - \kappa(a) \tag{B.41}$$

with $a > s$ has a central core region with a rising rotation curve for $\theta \lesssim s$, a flat rotation curve for $s \lesssim \theta \lesssim a$ and a dropping rotation curve for $\theta \gtrsim a$. In the singular limit ($s \to 0$), it becomes the "pseudo-Jaffe model" corresponding to a 3D density distribution $\rho \propto (r^2+s^2)^{-1}(r^2+a^2)^{-1}$ whose name derives from the fact that it is very similar the Jaffe model with $\rho \propto r^{-2}(r+a)^{-2}$ (Kneib et al. 1996, Keeton & Kochanek 1998). We will discuss other common lens models in §B.4.1.

The last simple analytic models we mention are the generalized singular isothermal potentials of the form $\Psi = \theta F(\chi)$ with surface density $\kappa(\theta,\chi) = (1/2)(F(\chi) + F''(\chi))/\theta$. Both the SIS and SIE are examples of this model. The generalized isothermal sphere has a number of useful analytic properties. For example, the magnification contours are isodensity contours

$$\mu^{-1} = 1 - \frac{1}{\theta}\left[F(\chi) + F''(\chi)\right] = 1 - 2\kappa(\theta,\chi) \tag{B.42}$$

with the tangential critical line being the contour with $\kappa = 1/2$, and the time delays between images depend only on the distances from the images to the lens center (see Witt, Mao & Keeton 2000, Kochanek, Keeton & McLeod 2001, Wucknitz 2002, Evans & Witt 2003).

B.4 The Mass Distributions of Galaxies

Contrary to popular belief, the modeling of gravitational lenses to determine the mass distribution of a lens is not a "black art." It is, however, an area in which the lensing community has communicated results badly. There are two main problems. First, many modeling results seem almost deliberately obfuscatory as to what models were actually used, what data were fit and what was actually constrained. Not only do many lens papers insist on taking well known density distributions from the dynamical literature and assigning them new names simply because they have been projected into two dimensions, but they then assign them a plethora of bizarre acronyms. Sometimes the model used is not actually the one named, for example using tidally truncated halos but calling them isothermal models. Second, there is a steady confusion between the parameters of models and the aspects of the mass distribution that have actually been constrained. Models with apparently very different parameters may be in perfect accord as to the properties of the mass distribution that are actually relevant to what is observed. Discussions of non-parametric mass models then confuse the issue further by conflating differences in parameters with differences in what is actually constrained to argue for non-parametric models when in fact they also are simply matching the same basic properties with lots of extra noise from the additional and uninteresting degrees of freedom. In short, the problem with lens modeling is not that it is a "black art," but that the practitioners try to make it seem to be a "black art" (presumably so that people will believe they need wizards). The most important point to take from this section is that any idiot can model a lens and interpret it properly with a little thinking about what it is that lenses constrain.

There are two issues to think about in estimating the mass distributions of gravitational lenses. The first issue is how to model the mass distribution with a basic choice between parametric and non-parametric models. In §B.4.1 we summarize the most commonly used radial mass distributions for lens models. Ellipsoidal versions of these profiles combined with an external (tidal) shear are usually used to describe the angular structure, but there has been recent interest in deviations from ellipsoidal distributions which we discuss in §B.4.4 and §B.8. In §B.4.7 we summarize the most common approaches for non-parametric models of the mass distribution. Since this is my review, I will argue that the parametric models are all that is needed to model lenses and that they provide a better basis for understanding the results than non-parametric models (but the reader should be warned that if Prasenjit Saha was writing this you would probably get a different opinion).

The second issue is to determine the aspects of the lens data that actually constrain the mass distribution. Among the things that can be measured for a lens are the relative positions of the components (the astrometric constraints), the relative fluxes of the images, the time delays between the images, the dynamical properties of the lens galaxy, and the microlensing of the images. Of

these, the most important constraints are the positions. We can usually measure the relative positions of the lensed components very accurately (5 mas or better) compared to the arc second scales of the component separations. Obviously the accuracy diminishes when components are faint, and the usual worst case is having very bright lensed quasars that make it difficult to detect the lens galaxy. As we discuss in §B.8, substructure and/or satellites of the lens galaxy set a lower limit of order 1–5 mas with which it is safe to impose astrometric constraints independent of the measurement accuracy. When the source is extended, the resulting arcs and rings discussed in §B.10 provide additional constraints. These are essentially astrometric in nature, but are considerably more difficult to use than multiply imaged point sources. Our general discussion of how lenses constrain the radial (§B.4.3) and angular structure (§B.4.4) focus on the use of astrometric constraints, and in §B.4.6 we discuss the practical details of fitting image positions in some detail.

The flux ratios of the images are one of the most easily measured constraints, but are cannot be imposed stringently enough to constrain radial density profiles because of systematic uncertainties. Flux ratios measured at a single epoch are affected by time variability in the source (§B.5), microlensing by the stars in the lens galaxy in the optical continuum (see Part 4), magnification perturbations from substructure at all wavelengths (see §B.8), absorption by the ISM of the lens (dust in the optical, free-free in the radio) and scatter broadening in the radio (see §B.8 and §B.9). Most applications of flux ratios have focused on using them to probe these perturbing effects rather than for studying the mean mass distribution of the lens. Where radio sources have small scale VLBI structures, the changes in the relative astrometry of the components can constrain the components of the relative magnification tensors without needing to use any flux information (e.g. Garrett et al. 1994, Rusin et al. 2002).

Two types of measurements, time delays (§B.5) and microlensing by the stars or other compact objects in the lens galaxy (Part 4) constrain the surface density near the lensed images. Microlensing also constrains the fraction of that surface density that can be in the form of stars. To date, time delays have primarily been used to estimate the Hubble constant rather than the surface density, but if we view the Hubble constant as a known quantity, consider only time delay ratios, or simply want to compare surface densities between lenses, then time delays can be used to constrain the mass distribution. We discuss time delays separately because of their close association with attempts to measure the Hubble constant. Using microlensing variability to constrain the mass distribution is presently more theory than practice due to a lack of microlensing light curves for almost all lenses. However, the light curves of the one well monitored lens, Q2337+0305, appear to require a surface density composed mainly of stars as we would expect for a lens where we see the images deep in the bulge of a nearby spiral galaxy (Kochanek 2004). We will not discuss this approach further in Part 2.

Any independent measurement of the mass of a component will also help to constrain the structure of the lenses. At present this primarily means making stellar dynamical measurements of the lens galaxy and comparing the dynamical mass estimates to those from the lens geometry. We discuss this in detail in §B.4.9. For lenses associated with clusters, X-ray, weak lensing or cluster velocity dispersion measurements can provide estimates of the cluster mass. While this has been done in a few systems (e.g. X-rays, Morgan et al. 2001, Chartas et al. 2002; weak lensing, Fischer et al. 1997; velocity dispersions, Angonin-Willaime, Soucail & Vanderriest 1994), the precision of these mass estimates is not high enough to give strong constraints on lens models. X-ray observations are probably more important for locating the positions of groups and clusters relative to the lens than for estimating their masses.

The most useful way of thinking about lensing constraints on mass distributions is in terms of multipole expansions (e.g. Kochanek 1991a, Trotter, Winn & Hewitt 2000, Evans & Witt 2003, Kochanek & Dalal 2004). An arbitrary surface density $\kappa(\boldsymbol{\theta})$ can be decomposed into multipole components,

$$\kappa(\boldsymbol{\theta}) = \kappa_0(\theta) + \sum_{m=1}^{\infty} \left[\kappa_{cm}(\theta)\cos(m\chi) + \kappa_{sm}(\theta)\sin(m\chi)\right] \tag{B.43}$$

where the individual components are angular averages over the surface density

$$\kappa_0(\theta) = \frac{1}{2\pi}\int_0^{2\pi} d\chi\kappa(\boldsymbol{\theta}), \quad \text{and} \quad \begin{pmatrix}\kappa_{cm}(\theta)\\ \kappa_{sm}(\theta)\end{pmatrix} = \frac{1}{\pi}\int_0^{2\pi} d\chi \begin{pmatrix}\kappa(\boldsymbol{\theta})cos(m\chi)\\ \kappa(\boldsymbol{\theta})\sin(m\chi)\end{pmatrix}. \tag{B.44}$$

The first three terms are the monopole (κ_0), the dipole ($m = 1$) and the quadrupole ($m = 2$) of the lens. The Poisson equation $\nabla^2\Psi = 2\kappa$ is separable in polar coordinates, so a multipole decomposition of the effective potential

$$\Psi(\boldsymbol{\theta}) = \Psi_0(\theta) + \sum_{m=1}^{\infty} \left[\Psi_{cm}(\theta)\cos(m\chi) + \Psi_{sm}(\theta)\sin(m\chi)\right] \tag{B.45}$$

will have terms that depend only on the corresponding multipole of the surface density, $\nabla^2\Psi_{cm}(\theta)\cos(m\chi) = 2\kappa_{cm}(\theta)\cos(m\chi)$. The monopole of the potential is simply

$$\Psi_0(\theta) = 2\log(\theta)\int_0^{\theta} udu\kappa_0(u) + 2\int_{\theta}^{\infty} udu\log(u)\kappa(u) \tag{B.46}$$

and its derivative is the bend angle for a circular lens,

$$\alpha_0(\theta) = \frac{d\Psi_0}{d\theta} = \frac{2}{\theta}\int_0^{\theta} udu\kappa_0(u), \tag{B.47}$$

just as we derived earlier (Eqn. B.3). The higher order multipoles are no more complicated, with

$$\begin{pmatrix}\Psi_{cm}(\theta)\\ \Psi_{sm}(\theta)\end{pmatrix} = -\frac{1}{m\theta^m}\int_0^\theta u^{1+m}du\begin{pmatrix}\kappa_{cm}(u)\\ \kappa_{sm}(u)\end{pmatrix} - \frac{\theta^m}{m}\int_\theta^\infty u^{1-m}du\begin{pmatrix}\kappa_{cm}(u)\\ \kappa_{sm}(u)\end{pmatrix}. \tag{B.48}$$

The angular multipoles are always composed of two parts. There is an interior pole $\Psi_{cm,int}(\theta)$ due to the multipole surface density interior to θ (the integral from $0 < u < \theta$) and an exterior pole $\Psi_{cm,ext}(\theta)$ due to the multipole surface density exterior to θ (the integral from $\theta < u < \infty$). The higher order multipoles produce deflections in both the radial

$$\alpha_{cm,rad} = \frac{d}{d\theta}\left[\Psi_{cm}\cos(m\chi)\right] = \frac{d\Psi_{cm}}{d\theta}\cos(m\chi) \tag{B.49}$$

and tangential

$$\alpha_{cm,tan} = \frac{1}{\theta}\frac{d}{d\chi}\left[\Psi_{cm}\cos(m\chi)\right] = -\frac{m}{\theta}\Psi_{cm}\sin(m\chi) \tag{B.50}$$

directions, where the radial deflection depends on the derivative of Ψ_{cm} and the tangential deflection depends only on Ψ_{cm}. This may seem rather formal, but the multipole expansion provides the basis for understanding which aspects of mass distributions will matter for lens models. Obviously it is the lowest order angular multipoles which are most important. The most common angular term added to lens models is the external shear

$$\Psi_{2,ext} = \frac{1}{2}\gamma_c\theta^2\cos 2(\chi-\chi_\gamma) + \frac{1}{2}\gamma_s\theta^2\sin 2(\chi-\chi_\gamma) \tag{B.51}$$

with dimensionless amplitudes γ_c and γ_s and axis χ_γ. The external (tidal) shear and any accompanying mean convergence are the lowest order perturbations from any object near the lens that have measurable effects on a gravitational lens (see Eqn. B.26). While models usually consider only external (tidal) shears where these coefficients are constants, in reality γ_c, γ_s and χ_γ are functions of radius (i.e. Eqn. B.48). Along with the external shear, there is an internal shear

$$\Psi_{2,int} = \frac{1}{2}\Gamma_c\frac{\langle\theta\rangle^4}{\theta^2}\cos 2(\chi-\chi_\Gamma) + \frac{1}{2}\Gamma_s\frac{\langle\theta\rangle^4}{\theta^2}\sin 2(\chi-\chi_\Gamma). \tag{B.52}$$

due to the quadrupole moment of the mass interior to a given radius. We introduce the mean radius of the lensed images $\langle\theta\rangle$ to make Γ_c and Γ_s dimensionless with magnitudes that can be easily compared to the external shear amplitudes γ_c and γ_s. Arguably the critical radius of the lens is a better physical choice, but the mean image radius will be close to the critical radius and using it avoids any trivial covariances between the internal shear strength and the monopole mass. Usually the internal quadrupole is added as part of an ellipsoidal model for the central lens galaxy, but it is useful in analytic studies to consider it separately.

B.4.1 Common Models for the Monopole

Most attention in modeling lenses focuses on the monopole or radial mass distribution of the lenses. Unfortunately, much of the lensing literature uses an almost impenetrable array of ghastly non-standard acronyms to describe the mass models even though many of them are identical to well-known families of density distributions used in stellar dynamics. Here we summarize the radial mass distributions which are most commonly used and will keep reappearing in the remainder of Part 2.

The simplest possible choice for the mass distribution is to simply trace the light. The standard model for early-type galaxies or the bulges of spiral galaxies is the de Vaucouleurs (1948) profile with surface density

$$\Sigma(R) = I_e \exp\left[-7.67\left[(R/R_e)^{1/4} - 1\right]\right], \tag{B.53}$$

where the effective radius R_e encompasses half the total mass (or light) of the profile. Although the central density of a de Vaucouleurs model is finite, it actually acts like a rather cuspy density distribution and will generally fit the early-type lens data with no risk of producing a detectable central image (e.g. Lehár et al. 2000, Keeton 2003a). The simplest model for a disk galaxy is an exponential disk,

$$\Sigma(R) = I_0 \exp\left[-R/R_d\right] \tag{B.54}$$

where R_d is the disk scale length. An exponential disk by itself is rarely a viable lens model because it has so little density contrast between the center and the typical radii of images that detectable central images are almost always predicted but not observed. Some additional component, either a de Vaucouleurs bulge or a cuspy dark matter halo, is always required. This makes spiral galaxy lens models difficult because they generically require two stellar components (a bulge and a disk) and a dark matter halo, while the photometric data are rarely good enough to constrain the two stellar components (e.g. Maller, Flores & Primack 1997, Koopmans et al. 1998, Maller et al. 2000, Trott & Webster 2002, Winn, Hall & Schechter 2003). Since spiral lenses are already relatively rare, and spiral lens galaxies with good photometry are rarer still, less attention has been given to these systems. The de Vaucouleurs and exponential disk models are examples of Sersic (1968) profiles

$$\Sigma(R) = I_0 \exp\left[-b_n\left[(R/R_e(n))^{1/n}\right]\right] \tag{B.55}$$

where the effective radius $R_e(n)$ is defined to encompass half the light and $n = 4$ is a de Vaucouleurs model and $n = 1$ is an exponential disk. These profiles have not been used as yet for the study of lenses except for some quasar host galaxy models (§B.10). The de Vaucouleurs model can be approximated (or the reverse) by the Hernquist (1990) model with the 3D density distribution

$$\rho(r) = \frac{M}{\pi r}\frac{a}{(a+r)^3} \tag{B.56}$$

and $a \simeq 0.55R_e$ if matched to a de Vaucouleurs model. For lensing purposes, the Hernquist model has one major problem. Its $\rho \propto 1/r$ central density cusp is shallower than the effective cusp of a de Vaucouleurs model, so Hernquist models tend to predict detectable central images even when the matching de Vaucouleurs model would not. As a result, the Hernquist model is more often used as a surrogate for dynamical normalization of the de Vaucouleurs model than as an actual lens model (see below).

Theoretical models for lenses started with simple, softened power laws of the form

$$\kappa(R) \propto \left(R^2 + s^2\right)^{-(n-1)/2} \to R^{1-n} \tag{B.57}$$

in the limit where there is no core radius. We are using these simple power law lenses in all our examples (see §B.3). These models include many well known stellar dynamical models such as the singular isothermal sphere (SIS, $n = 2$, $s = 0$), the modified Hubble profile ($n = 3$) and the Plummer model ($n = 5$). Since we only see the projected mass, these power laws are also related to common models for infinitely thin disks. The Mestel (1963) disk ($n = 2$, $s = 0$) is the disk that produces a flat rotation curve, and the Kuzmin (1956) disk ($n = 3$) can be used to mimic the rising and then falling rotation curve of an exponential disk. The softened power-law models have generally fallen out of favor other than as simple models for some of the visible components of lenses because the strong evidence for stellar and dark matter cusps makes models with core radii physically unrealistic. While ellipsoidal versions of these models are not available in useful form, there are fast series expansion methods for numerical models (Chae, Khersonsky & Turnshek 1998, Barkana 1998).

Most "modern" discussions of galaxy density distributions are based on sub-cases of the density distribution

$$\rho(r) \propto \frac{1}{r^n} \frac{1}{\left(a^\alpha + r^\alpha\right)^{(m-n)/\alpha}}, \tag{B.58}$$

which has a central density cusp with $\rho \propto r^{-n}$, asymptotically declines as $\rho \propto r^{-m}$ and has a break in the profile near $r \simeq a$ whose shape depends on α (e.g. Zhao 1997). The most common cases are the Hernquist model ($n = 1$, $m = 4$, $\alpha = 1$) mentioned above, the Jaffe (1983) model ($n = 2$, $m = 4$, $\alpha = 1$), the NFW (Navarro, Frenk & White 1996) model ($n = 1$, $m = 3$, $\alpha = 1$) and the Moore (1998) model ($n = 3/2$, $m = 3$, $\alpha = 1$). We can view the power-law models either as the limit $n \to 0$ and $\alpha = 2$, or we could generalize the r^{-n} term to $(r^2 + s^2)^{-n/2}$ and consider only regions with r and $s \ll a$. Projections of these models are similar to surface density distributions of the form

$$\kappa(R) \propto \frac{1}{R^{n-1}} \frac{1}{\left(a^\alpha + R^\alpha\right)^{(m-n)/\alpha}} \tag{B.59}$$

(although the definition of the break radius a changes) with the exception of the limit $n \to 1$ where the projection of a 3D density cusp $\rho \propto 1/r$ produces

surface density terms $\kappa \propto \ln R$ that cannot be reproduced by the broken surface density power law. This surface density model is sometimes called the Nuker law (e.g. Byun et al. 1996). A particularly useful case for lensing is the pseudo-Jaffe model with $n = 2, m = 4$ and $\alpha = 2$ (where the normal Jaffe model has $\alpha = 1$) as the only example of a broken power law with simple analytic deflections even when ellipsoidal because the density distribution is the difference between two isothermal ellipsoids (see Eqn. B.41). These cuspy models also allow fast approximate solutions for their ellipsoidal counterparts (see Chae 2002).

The most theoretically important of these cusped profiles is the NFW profile (Navarro et al. 1996) because it is the standard model for dark matter halos. Since it is such a common model, it is worth discussing it in a little more detail, particularly its peculiar normalization. The NFW profile is normalized by the mass M_{vir} inside the virial radius r_{vir}, with

$$\rho_{NFW}(r) = \frac{M_{vir}}{4\pi f(c)} \frac{1}{r(r+a)^2} \quad \text{and} \quad M_{NFW}(< r) = \frac{f(r/r_{vir})}{f(c)} \tag{B.60}$$

where $f(c) = \ln(1+c) - c/(1+c)$ and the concentration $c = r_{vir}/a \sim 5$ for clusters and $c \sim 10$ for galaxies. The concentration is a function of mass whose scaling is determined from N-body simulations. A typical scaling for a halo at redshift z in an $\Omega_M = 0.3$ flat cosmological models is (Bullock et al. 2001)

$$c(M) = \frac{9}{1+z} \left(\frac{M_{vir}}{8 \times 10^{12} h M_\odot} \right)^{-0.14} \tag{B.61}$$

with a dispersion in $\log c$ of $\sigma_{\log(c)} \simeq 0.18$ dex. Because gravitational lensing is very sensitive to the central density of the lens, including the scatter in the concentration is quantitatively important for lensing by NFW halos (Keeton 2001b). The virial mass and radius are related and determined by the overdensity $\Delta_{vir}(z)$ required for a halo to collapse given the cosmological model and the redshift. This can be approximated by

$$M_{vir} = \frac{4\pi}{3} \Delta_{vir}(z) \rho_u(z) r_{vir}^3 \simeq 0.23 \times 10^{12} h \left(\frac{(1+z) r_{vir}}{100 h^{-1} \mathrm{kpc}} \right)^3 \left(\frac{\Omega_M \Delta_{vir}}{200} \right) M_\odot \tag{B.62}$$

where $\rho_u(z) = 3H_0^2 \Omega_M (1+z)^3 / 8\pi G$ is the mean matter density when the halo forms and $\Delta_{vir} \simeq (18\pi^2 + 82x - 39x^2)/\Omega(z)$ with $x = \Omega - 1$ is the overdensity needed for a halo to collapse. There are differences in normalizations between authors and with changes in the central cusp exponent γ, but models of this type are what we presently expect for the structure of dark matter halos around galaxies.

For most lenses, HST imaging allows us to measure the spatial distribution of the stars, thereby providing us with a model for the distribution of stellar mass with only the stellar mass-to-light ratio as a parameter. For present purposes, gradients in the stellar mass-to-light ratio are unimportant compared

to the uncertainties arising from the dark matter. Unless we are prepared to abandon the entire paradigm for modern cosmology, the luminous galaxy is embedded in a dark matter halo and we must decide how to model the overall mass distribution. The most common approach, as suggested by the rich variety of mass profiles we introduced in §B.4.1, is to assume a parametric form for the total mass distribution rather than attempting to decompose it into luminous and dark components. The alternative is to try to embed the stellar component in a dark matter halo. Operationally, doing so is trivial – the lens is simply modeled as the sum of two mass components. However, there are theoretical models for how CDM halos should be combined with the stellar component.

Most non-gravitational lensing applications focus on embedding disk galaxies in halos because angular momentum conservation provides a means of estimating a baryonic scale length (e.g. Mo, Mao & White 1998). The spin parameter of the halo sets the angular momentum of the baryons, and the final disk galaxy is defined by the exponential disk with the same angular momentum. As the baryons become more centrally concentrated, they pull the dark matter inwards as well through a process known as adiabatic contraction (Blumenthal et al. 1986). The advantage of this approach, which in lensing has been used only by Kochanek & White (2001), is that it allows a full *ab initio* calculation of lens statistical properties when combined with a model for the cooling of the baryons (see §B.7). It has the major disadvantage that most lens galaxies are early-type galaxies rather than spirals, and that there is no analog of the spin parameter and angular momentum conservation to set the scale length of the stellar component in a model for an early-type galaxy.

Models of early-type galaxies embedded in CDM halos have to start with an empirical estimate of the stellar effective radius. In models of individual lenses this is a measured property of the lens galaxy (e.g. Rusin et al. 2003, 2004, Koopmans & Treu 2002, Kochanek 2003a). Statistical models must use a model for the scaling of the effective radius with luminosity or other observable parameters of early-type galaxies (e.g. Keeton 2001b). From the luminosity, a mass-to-light ratio is used to estimate the stellar mass. If all baryons have cooled and been turned into stars, then the stellar mass provides the total baryonic mass of the halo, otherwise the stellar mass sets a lower bound on the baryonic mass. Combining the baryonic mass with an estimate of the baryonic mass fraction yields the total halo mass to be fed into the model for the CDM halo.

In general, there is no convincing evidence favoring either approach – for the regions over which the mass distributions are constrained by the data, both approaches will agree on the overall mass distribution. However, there can be broad degeneracies in how the total mass distribution is decomposed into luminous and dark components (see §B.4.6).

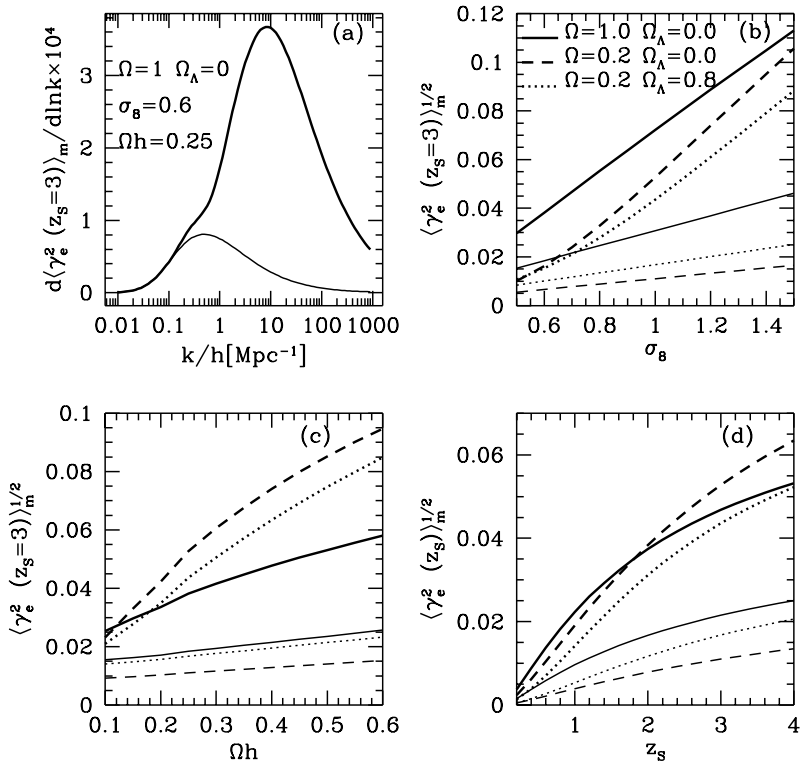


Fig. B.19. Dependence of the shear generated by other objects along the line of sight for both linear (light lines) and non-linear (heavy lines) power spectra. (a) Shows the logarithmic contribution to the rms effective shear for a source at redshift $z_s = 3$ as a function of wave vector k. (b) Shows the dependence on σ_8 for a fixed power spectrum shape $\Omega_M h = 0.25$. (c) Shows the dependence on the shape $\Omega_M h$ with $\sigma_8 = 0.6$ for $\Omega_M = 1$ and $\sigma_8 = 1.0$ for $\Omega_M < 1$. (d) Shows the variation in the shear with source redshift for the models in (c) with $\Omega_M h = 0.25$.

B.4.2 The Effective Single Screen Lens

Throughout these notes we will treat lenses as if all the lens components lay at a single redshift ("the single screen approximation"). The lens equations for handling multiple deflection screens (e.g. Blandford & Narayan 1986, Kovner 1987b, Barkana 1996) are known but little used except for numerical studies (e.g. Kochanek & Apostolakis 1998, Moller & Blain 2001) in large part because few lenses require multiple lens galaxies at different redshifts with the exception of B2114+022 (Chae, Mao & Augusto 2001). In fact, we are not being as cavalier in making this approximation as it may seem.

The vast majority of strong lenses consist of a single lens galaxy perturbed by other objects. We can divide these objects into those near the primary lens, where a single screen is clearly appropriate, and those distributed along the line of site for which a single screen may be inappropriate. Because the correlation function is so strong on small scales, the perturbations are dominated by objects within a correlation length of the lens galaxy (e.g. Keeton, Kochanek & Seljak 1997, Holder & Schechter 2003). The key to the relative safety of the single screen model is that weak perturbations from objects along the line of site, in the sense that in a multi-screen lens model they could be treated as a convergence and a shear, can be reduced to a single "effective" lens plane in which the true amplitudes of the convergence and shear are rescaled by distance ratios to convert them from their true redshifts to the redshift of the single screen (Kovner 1987b, Barkana 1996). The lens equation on the effective single screen takes the form

$$\boldsymbol{\beta} = (I + F_{OS})\,\boldsymbol{\theta} - (I + F_{LS})\,\boldsymbol{\alpha}\left[(I + F_{OL})\,\boldsymbol{\theta}\right] \tag{B.63}$$

where F_{OS}, F_{LS} and F_{OL} describe the shear and convergence due to perturbations between the observer and the source, the lens and the source and the observer and the lens respectively. For statistical calculations this can be simplified still further by making the coordinate transformation $\boldsymbol{\theta}' = (I + F_{OL})\boldsymbol{\theta}$ and $\boldsymbol{\beta}' = (I + F_{LS})\boldsymbol{\beta}$) to leave a lens equation,

$$\boldsymbol{\beta}' = (I + F_e)\,\boldsymbol{\theta}' - \boldsymbol{\alpha}\left[\boldsymbol{\theta}'\right], \tag{B.64}$$

identical to a single screen lens in an effective convergence and shear of $F_e = F_{OL} + F_{LS} - F_{OS}$ (to linear order). In practice it will usually be safe to neglect the differences between Eqns. B.63 and B.64 because the shearing terms affecting the deflections in Eqn B.63 are easily mimicked by modest changes in the ellipticity and orientation of the primary lens. The rms amplitudes of these perturbations depend on the cosmological model and the amplitude of the non-linear power spectrum, but the general scaling is that the perturbations grow as $D_s^{3/2}$ with source redshift, and increase for larger σ_8 and Ω_M as shown in Fig. B.19 from Keeton et al. (1997). The importance of these effects is very similar to concerns about the effects of lenses along the line of sight on the brightness of high redshift supernova being used to estimate the cosmological model (e.g. Dalal et al. 2003).

B.4.3 Constraining the Monopole

The most frustrating aspect of lens modeling is that it is very difficult to constrain the monopole. If we take a simple lens and fit it with any of the parametric models from §B.4.1, it will be possible to obtain a good fit provided the central surface density of the model is high enough to avoid the formation of a central image. As usual, it is simplest to begin understanding

the problem with a circular, two-image lens whose images lie at radii θ_A and θ_B from the lens center (Fig. B.20). The lens equation (B.4) constrains the deflections so that the two images correspond to the same source position,

$$\beta = \theta_A - \alpha(\theta_A) = -\theta_B + \alpha(\theta_B), \tag{B.65}$$

where the sign changes appear because the images are on opposite sides of the lens. Recall that for the power-law lens model, $\alpha(\theta) = b^{n-1}\theta^{2-n}$ (Eqn. B.9), so we can easily solve the constraint equation to determine the Einstein radius of the lens,

$$b = \left[\frac{\theta_A + \theta_B}{\theta_A^{2-n} + \theta_B^{2-n}}\right]^{1/(n-1)} \tag{B.66}$$

in terms of the image positions. In the limit of an SIS ($n = 2$) the Einstein radius is the arithmetic mean, $b = (\theta_A + \theta_B)/2$, and in the limit of a point source ($n \to 3$), it is the geometric mean, $b = (\theta_A\theta_B)^{1/2}$, of the image radii. More generally, for any deflection profile $\alpha(\theta) = bf(\theta)$, the two images simply determine the mass scale $b = (\theta_A + \theta_B)/(f(\theta_A) + f(\theta_B))$.

There are two important lessons here. First, the location of the tangential critical line is determined fairly accurately independent of the mass profile. We may only be able to determine the mass scale, but it is the most accurate measurement of galaxy masses available to astronomy. The dependence of the mass inside the Einstein radius on the shape of the deflection profile is weak, with fractional differences between profiles being of order $(\delta\theta/\langle\theta\rangle)^2/8$ where $\delta\theta = \theta_A - \theta_B$ and $\langle\theta\rangle = (\theta_A + \theta_B)/2$ (i.e. if the images have similar radii, the difference beween the arthmetic and geometric mean is small). Second, it is going to be very difficult to determine radial mass distributions. In this example there is a perfect degeneracy between the exact location of the tangential critical line b and the exponent n. In theory, this is broken by the flux ratio of the images. However, a simple two-image lens has too few constraints even with perfectly measured flux ratios because a realistic lens model must also include some freedom in the angular structure of the lens. For a simple four-image lens, there begin to be enough constraints but the images all have similar radii, making the flux ratios relatively insensitive to changes in the monopole. Combined with the systematic uncertainties in flux ratios, they are not useful for this purpose.

This example also leads to the major misapprehension about lens models and radial mass distributions, in that the constraints appear to lead to a degeneracy related to the global structure of the potential (i.e. the exponent n). This is not correct. The degeneracy is a purely local one that depends only on the structure of the lens in the annulus defined by the images, $\theta_B < \theta < \theta_A$, as shown in Fig. B.20. To see this we will rewrite the expression for the bend

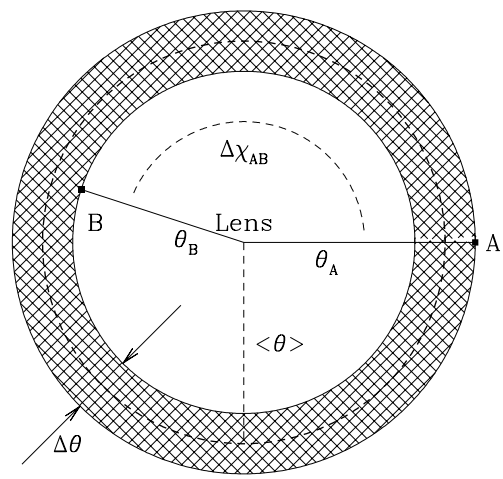


Fig. B.20. A schematic diagram of a two-image lens. The lens galaxy lies at the origin with two images A and B at radii θ_A and θ_B from the lens center. The images define an annulus of average radius $\langle\theta\rangle = (\theta_A + \theta_B)/2$ and width $\delta\theta = \theta_A - \theta_B$, and they subtend an angle $\Delta\chi_{AB}$ relative to the lens center. For a circular lens $\Delta\chi_{AB} = 180°$ by symmetry.

angle (Eqn. B.3) as

$$\alpha(\theta) = \frac{2}{\theta}\left[\int_0^{\theta_B} u du \kappa(u) + \int_{\theta_B}^{\theta} u du \kappa(u)\right] = \frac{1}{\theta}\left[b_B^2 + (\theta^2 - \theta_B^2)\langle\kappa\rangle(\theta, \theta_B)\right] \tag{B.67}$$

where $b_B^2 = 2\int_0^{\theta_B} u du \kappa(u)$ is the Einstein radius of the total mass interior to image B, and

$$\langle\kappa\rangle(\theta, \theta_B) = \frac{2}{\theta^2 - \theta_B^2}\int_{\theta_B}^{\theta} u du \kappa(u) \tag{B.68}$$

is the mean surface density in the annulus $\theta_B < u < \theta$. If we now solve the constraint Eqn. B.65 again, we find that

$$b_B^2 = \theta_A\theta_B - \langle\kappa\rangle_{AB}\theta_B(\theta_A - \theta_B) \tag{B.69}$$

where $\langle\kappa\rangle_{AB} = \langle\kappa\rangle(\theta_A, \theta_B)$ is the mean density in the annulus $\theta_B < \theta < \theta_A$ between the images. Thus, there is a degeneracy between the total mass interior to image B and the mean surface density (mass) between the two images. There is no dependence on the distribution of the mass interior to

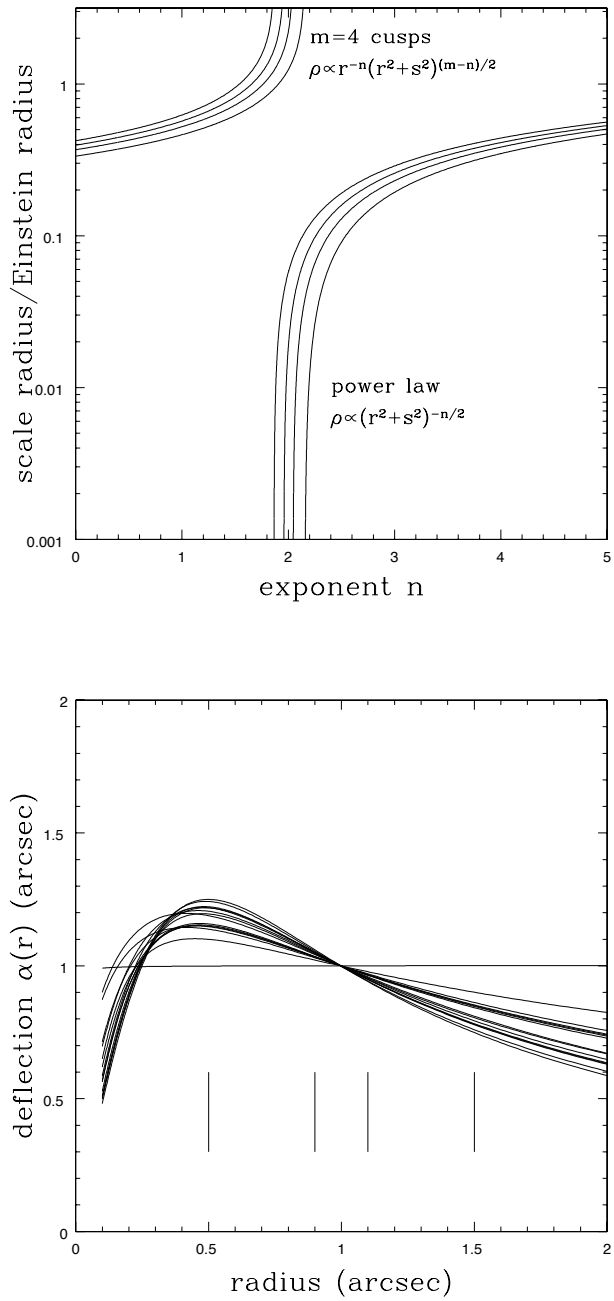


Fig. B.21. Softened power law and cusped model fits to the images produced by an SIS lens with Einstein radius $b = 1\farcs0$ and two source components located $0\farcs1$ and $0\farcs5$ from the lens center. In the top panel, the contours show the regions with astrometric fit residuals per image of $0\farcs003$ and $0\farcs010$. Models with $m = 3$ cusps so closely overly the $m = 4$ models that their error contours were not plotted. The bottom panel shows the deflection profiles of the best models at half-integer increments in the exponent n. The SIS model has a constant deflection, and the power-law and cusp models approach it in a sequence of slowly falling deflection profiles. All models agree with the SIS Einstein radius at $r = 1\farcs0$. The positions of the images are indicated by the vertical bars.

θ_B, the distribution of mass between the two images, or on either the amount or distribution of mass exterior to θ_A. This is Gauss' law for gravitational lens models.

If we normalize the mass scale at any point in the interior of the annulus then the result will appear to depend on the distribution of the mass simply because the mass must be artificially divided. For example, suppose we model the surface density *locally* as a power law $\kappa \propto \theta^{1-n}$ with a mean surface density $\langle\kappa\rangle$ in the annulus $\theta_B < \theta < \theta_A$ between the images. The mass inside the mean image radius $\langle\theta\rangle$ is

$$b^2_{\langle\theta\rangle} = \theta_A\theta_B\,(1-\kappa_0) + \delta\theta^2\langle\kappa\rangle\left[\frac{n}{4} + \left(\frac{\delta\theta}{\langle\theta\rangle}\right)^2 \frac{(4-n)(2-n)(1-n)}{192} + O\left(\left(\frac{\delta\theta}{\langle\theta\rangle}\right)^4\right)\right] \tag{B.70}$$

where we have expanded the result in the ratio $\delta\theta/\langle\theta\rangle$ (in fact, the result as shown is exact for $n = 2/3$, 1, 2, 4 and 5). We included in this result an additional, global convergence κ_0 so that we can contrast the local degeneracies due to the distribution of matter between the images with the global degeneracies produced by a infinite mass sheet. The leading term $\theta_A\theta_B$ is the Einstein radius expected for a point mass lens (Eqn. B.65). While the total enclosed mass ($\theta_A\theta_B$) is fixed, the mass associated with the lens galaxy $b^2_{\langle\theta\rangle}$ must be modified in the presence of a global convergence by the usual $1-\kappa_0$ factor created by the mass sheet degeneracy (Falco, Gorenstein & Shapiro 1985). The structure of the lens in the annulus leads to fractional corrections to the mass of order $(\delta\theta/\langle\theta\rangle)^2$ that are proportional to $n\langle\kappa\rangle$ to lowest order.

Only if you have additional images inside the annulus can you begin to constrain the structure of the density in the annulus. The constraint is not, unfortunately, a simple constraint on the density. Suppose that we see an additional (pair) of images on the Einstein ring at θ_0, with $\theta_B < \theta_0 < \theta_A$ This case is simpler than the general case because it divides our annulus into two sub-annuli (from θ_B to θ_0 and from θ_0 to θ_A) rather than three. Since we put the extra image on the Einstein ring, we know that the mean surface density interior to θ_0 is unity (Eqn. B.11). The A and B images then constrain a ratio

$$\frac{1-\langle\kappa\rangle_{B0}}{1-\langle\kappa\rangle_{A0}} = \frac{\theta_B}{\theta_A}\frac{\theta_A^2-\theta_0^2}{\theta_0^2-\theta_B^2} \simeq \frac{\theta_A-\theta_0}{\theta_0-\theta_B}\left[1-\frac{\theta_A-\theta_B}{2\theta_0}\cdots\right] \tag{B.71}$$

of the average surface densities between the Einstein ring and image B ($\langle\kappa\rangle_{B0}$) and the Einstein ring and image A ($\langle\kappa\rangle_{A0}$). Since a physical distribution must have $0 < \langle\kappa\rangle_{A0} < \langle\kappa\rangle_{B0}$, the surface density in the inner sub-annulus must satisfy

$$\frac{\theta_A+\theta_B}{\theta_A}\frac{\theta_0^2-\theta_A\theta_B}{\theta_0^2-\theta_B^2} < \langle\kappa\rangle_{B0} < 1 \tag{B.72}$$

where the lower (upper) bound is found when the density in the outer sub-annulus is zero (when $\langle\kappa\rangle_{B0} = \langle\kappa\rangle_{A0}$). The term $\theta_0^2 - \theta_A\theta_B$ is the difference between the measured critical radius θ_0 and the critical radius implied by the other two images for a lens with no density in the annulus (e.g. a point mass), $(\theta_A\theta_B)^{1/2}$. Suppose we actually have images formed by an SIS, so $\theta_A = \theta_0(1+x)$ and $\theta_B = \theta_0(1-x)$ with $0 < x = \beta/\theta_0 < 1$, then the lower bound on the density in the inner sub-annulus is

$$\langle\kappa\rangle_{B0} > \frac{2x}{(2-x)(1+x)} \tag{B.73}$$

and the fractional uncertainly in the surface density is unity for images near the Einstein ring ($x \to 0$) and then steadily diminishes as the A and B images are more asymmetric. If you want to constrain the monopole, the more asymmetric the configuration the better. This rule becomes still more important with the introduction of angular structure.

Fig. B.21 illustrates these issues. We arbitrarily picked a model consisting of an SIS lens with two sources. One source is close to the origin and produces images at $\theta_A = 1\farcs1$ and $\theta_B = 0\farcs9$. The other source is farther from the origin with images at $\theta_A = 1\farcs5$ and $\theta_B = 0\farcs5$. We then modeled the lens with either a softened power law (Eqn. B.57) or a three-dimensional cusp (Eqn. B.58). We did not worry about the formation of additional images when the core radius becomes too large or the central cusp is too shallow – this would rule out models with very large core radii or shallow central cusps. If there were only a single source, either of these models can fit the data for any values of the parameters. Once, however, there are two sources, most of parameter space is ruled out except for degenerate tracks that look very different for the two mass models. Along these tracks, the models satisfy the additional constraint on the surface density given by Eqn. B.71. The first point to make about Fig. B.21 is the importance of carefully defining parameters. The input SIS model has very different parameters for the two mass models – while the exponent $n = 2$ is the same in both cases, the SIS model is the limit $s \to 0$ for the core radius in the softened power law, but it is the limit $a \to \infty$ for the break radius in the cusp model. Similarly, models with an inner cusp $n = 0$ will closely resemble power law models whose exponent n matches the outer exponent m of the cuspy models. Our frequent failure to explain these similarities is one reason why lens modeling seems so confusing. The second point to make about Fig. B.21 is that the deflection profiles implied by these models are fairly similar over the annulus bounded by the images. Outside the annulus, particularly at smaller radii, they start to show very large fractional differences. Only if we were to add a third set of multiple images or measure a time delay with a known value of H_0 would the parameter degeneracy begin to be broken.

These general results show that studies of how lenses constrain the monopole need the ability to simultaneously vary the mass scale, the surface density of the annulus and possibly the slope of the density profile in the annulus to

have the full range of freedom permitted by the data. Most parametric studies constraining the monopole have had two parameters, adjusting the mass scale and a correlated combination of the surface density and slope (e.g. Kochanek 1995a, Impey et al. 1998, Chae, Turnshek & Khersonsky 1998, Barkana et al. 1999, Chae 1999, Cohn et al. 2001, Muñoz et al. 2001, Wucknitz et al. 2004), although there are exceptions using models with additional degrees of freedom (e.g. Bernstein & Fischer 1999, Keeton et al. 2000, Trott & Webster 2002, Winn, Rusin & Kochanek 2003). This limitation is probably not a major handicap, because realistic density profiles show a rather limited range of local logarithmic slopes.

B.4.4 The Angular Structure of Lenses

Assuming you have identified all the halos needed to model a particular lens, there are three sources of angular structure in the potential. The first source is the shape of the luminous lens galaxy, the second source is the dark matter in the halo of the lens, and the third source is perturbations from nearby objects or objects along the line of sight. Of these, the only one which is easily normalized is the contribution from the stars in the lens galaxy, since it must be tightly connected to the monopole deflection of the stars. The observed axis ratios of early-type galaxies show a deficit of round galaxies, a plateau for axis ratios from $q \sim 0.9$ to $q \sim 0.5$ and then a sharp decline beyond $q \sim 0.5$ (e.g. Khairul & Ryden 2002). Not surprisingly, the true elliptical galaxies are rounder than the lenticular (S0) galaxies even if both are grouped together as early-type galaxies. In three dimensions, the stellar distributions are probably close to oblate with very modest triaxialities (e.g. Franx et al. 1991). Theoretical models of galaxy formation predict ellipticities and triaxialities larger than observed for luminous galaxies and show that the shapes of the dark matter halos are significantly modified by the cooling baryons (Dubinski 1992, 1994, Warren et al. 1992, Kazantzidis et al. 2004). Local estimates of the shape of dark matter halos are very limited (e.g. Olling & Merrifield 2001, Buote et al. 2002). Stellar isophotes also show deviations from perfect ellipses (e.g. Bender et al. 1989, Rest et al. 2001) and the deviations of simulated halos from ellipses have a similar amplitude (Heyl et al. 1994, Burkert & Naab 2003).

It is worth considering two examples to understand the relative importance of the higher order multipoles of a lens. The first is the singular isothermal ellipsoid (SIE) introduced in §B.3 (Eqns. B.38-B.40). Let the major axis of the model lie on the θ_1 axis, in which case only the $\cos(m\chi)$ multipoles with $m = 2, 4, \cdots$ are non-zero. All non-zero poles also have the same radial dependence, with $\kappa_{cm} = A_m/\theta$ and $\Psi_{cm} = -2A_m\theta/(m^2-1)$. The ratio of the internal to the external multipole depends only on the index of the multipole, $\Psi_{cm,int}/\Psi_{cm,ext} = (m-1)/(m+1)$. Note, in particular, that the quadrupole moment of an SIE is dominated by the matter *outside* any given radius, with

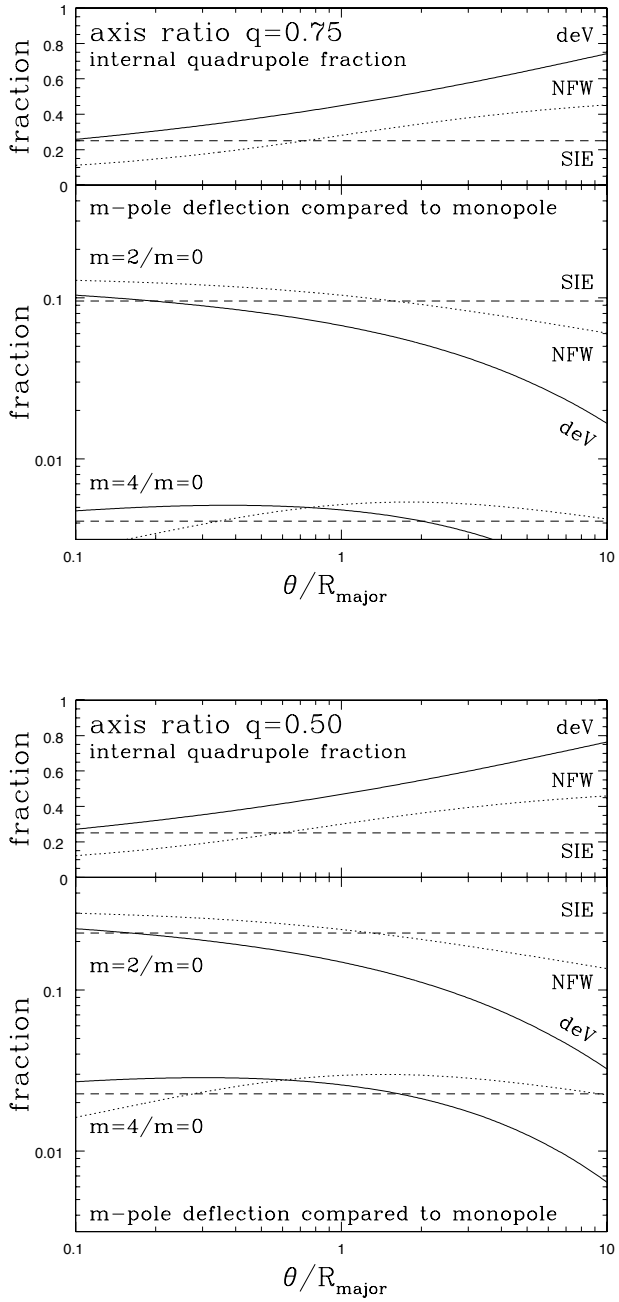


Fig. B.22. Behavior of the angular multipoles for the de Vaucouleurs (solid), SIE (dashed) and NFW (dotted) models with axis ratios of either $q = 0.75$ (Top) or $q = 0.5$ (Bottom) as a function of radius from the lens center in units of the lens major axis scale R_{major}. For each axis ratio, the lower panel shows the ratio of the maximum angular deflections produced by the quadrupole ($m = 2$) and the $m = 4$ pole relative to the deflection produced by the monopole ($m = 0$). The upper panel shows the fraction of the quadrupole generated by the mass interior to each radius.

an internal quadrupole fraction of

$$f_{int} = \frac{\Psi_{c2,int}}{\Psi_{c2}} = \frac{1}{4}. \tag{B.74}$$

For lenses dominated by dark matter halos that have roughly flat global rotation curves, most of the quadrupole moment is generated outside the Einstein ring of the lens (i.e. by the halo!). This will hold provided any halo truncation radius is large compared to the Einstein ring radius. The tangential deflection is larger than the radial deflection, with $|\alpha_{cm,rad}/\alpha_{cm,tan}| = 1/m$. The final question is the relative amplitudes between the poles. The ratio of the angular deflection from the $m = 2$ quadrupole to the radial deflection of the monopole is

$$\frac{\alpha_{c2,tan}}{\alpha_{0,rad}} \simeq \frac{\epsilon}{3}\left[1 + \frac{1}{2}\epsilon + \frac{9}{32}\epsilon^2 \cdots\right] \tag{B.75}$$

while the ratio for the $m = 4$ quadrupole is

$$\frac{\alpha_{c4,tan}}{\alpha_{0,rad}} \simeq \frac{\epsilon^2}{20}\left[1 + \epsilon + \frac{19}{24}\epsilon^2 \cdots\right] \tag{B.76}$$

where the axis ratio of the ellipsoid is $q = 1 - \epsilon$. Each higher order multipole has an amplitude $\Psi_m \propto \epsilon^{m/2}$ to leading order.

The relative importance of the higher order poles can be assessed by computing the deflections for a typical lens with the monopole deflection (essentially the Einstein radius) fixed to be one arc second. Using the leading order scaling of the power-series, but setting the numerical value to be exact for an axis ratio $q = 1/2$, the angular deflection from the quadrupole is $0\farcs46\epsilon$ and that from the $m = 4$ pole is $0\farcs09\epsilon^2$, while the radial deflections will be smaller by a factors of 2 and 4 respectively. Since typical astrometric errors are of order $0\farcs005$, the quadrupole is quantitatively important for essentially any ellipticity while the $m = 4$ pole becomes quantitatively important only for $q \lesssim 0.75$ (and the $m = 6$ pole becomes quantitatively important for $q \lesssim 0.50$).

In Fig. B.22 we compare the SIE to ellipsoidal de Vaucouleurs and NFW models. Unlike the SIE, these models are not scale free, so the multipoles depend on the distance from the lens center in units of the major axis scale length of the lens, R_{major}. The behavior of the de Vaucouleurs model will be typical of any ellipsoidal mass distribution that is more centrally concentrated than an SIE. Although the de Vaucouleurs model produces angular deflections similar to those of an SIE on small scales (for the same axis ratio), these are beginning to decay rapidly at the radii where we see lensed images (1–$2R_{major}$) because most of the mass is interior to the image positions and the amplitudes of the higher order multipoles decay faster with radius than the monopole (see Eqn. B.48). Similarly, as more of the mass lies at smaller radii, the quadrupole becomes dominated by the internal quadrupole. The NFW model has a somewhat different behavior because on small scales it is less

centrally concentrated than an SIE (a $\rho \propto 1/r$ central density cusp rather than $\propto 1/r^2$). It produces a somewhat bigger quadrupole for a given axis ratio, and an even larger fraction of that quadrupole is generated on large scales. In a "standard" dark matter halo model, the region with $\theta < R_{major}$ is also where we see the lensed images. On larger scales, the NFW profile is more centrally concentrated than the SIE, so the quadrupole begins to decay and becomes dominated by the internal component.

It is unlikely that mass distributions are true ellipsoids producing only even poles ($m = 2, 4, \cdots$) with no twisting of the axes with radius. For model fits we need to consider the likely amplitude of these deviations and the ability of standard terms to absorb and mask their presence. It is clear from Fig. B.22 that the amplitude of any additional terms must be of order the $m = 4$ deflections expected for an ellipsoid for them to be important. Here we illustrate the issues with the first few possible terms.

A dipole moment ($m = 1$) corresponds to making the galaxy lopsided with more mass on one side of the lens center than the other. Lopsidedness is not rare in disk galaxies ($\sim$30% at large radii, Zaritsky & Rix 1997), but is little discussed (and hence presumably small) for early-type galaxies. Certainly in the CASTLES photometry of lens galaxies we never see significant dipole residuals. It is difficult (impossible) to have an equilibrium system supported by random stellar motions with a dipole moment because the resulting forces will tend to eliminate the dipole. Similar considerations make it difficult to have a dark matter halo offset from the luminous galaxy. Only disks, which are supported by ordered rather than random motion, permit relatively long-lived lopsided structures. Where a small dipole exists, it will have little effect on the lens models unless the position of the lens galaxy is imposed as a stringent constraint. The reason is that a dipole adds terms to the effective potential of the form $\theta_1 G(\theta)$ whose leading terms are degenerate with a change in the unknown source position.

Perturbations to the quadrupole (relative to an ellipsoid) arise from variations in the ellipticity or axis ratio with radius. Since realistic lens models require an independent external shear simply to model the local environment, it will generally be very difficult to detect these types of perturbations or for these types of perturbations to significantly modify any conclusions. In essence, the amplitude and orientation of the external shear can capture most of their effects. Their actual amplitude is easily derived from perturbation theory. For example, if there is an isophote twist of $\Delta\chi$ between the region inside the Einstein ring and outside the Einstein ring, the fractional perturbations to the quadrupole will be of order $\Delta\chi$, or approximately $\epsilon\Delta\chi/3$ of the monopole – independent of the ability of the external shear to mimic the twist, the actual amplitude of the perturbation is approaching the typical measurement precision unless the twist is very large. Only in Q0957+561 have models found reasonably clear evidence for an effect arising from isophotal twists and ellipticity gradients, but both distortions are unusually large in

this system (Keeton et al. 2000). In general, in the CASTLES photometry of lens galaxies, deviations from simple ellipsoidal models are rare.

Locally we observe that the isophotes of elliptical galaxies are not perfect ellipses (e.g. Bender et al. 1989, Rest et al. 2001) and simulated halos show deviations of similar amplitude (Heyl et al. 1994, Burkert & Naab 2003). For lensing calculations it is useful to characterize these perturbations by a contribution to the lens potential and surface density of

$$\Psi = \frac{\epsilon_m}{m}\theta \cos m(\chi - \chi_m) \quad \text{and} \quad \kappa_m = \frac{\epsilon_m}{\theta}\frac{1-m^2}{m}\cos m(\chi - \chi_m) \tag{B.77}$$

respectively where the amplitude of the term is related to the usual isophote parameter $a_m = \epsilon_m|1-m^2|/mb$ for a lens with Einstein radius b. A typical early-type galaxy might have $|a_4| \sim 0.01$, so their fractional effect on the deflections, $|\epsilon_4|/b \sim |a_4|/4 \sim 0.003$, will be comparable to the astrometric measurement accuracy.

B.4.5 Constraining Angular Structure

The angular structure of lenses is usually simply viewed as an obstacle to understanding the monopole. This is a serious mistake. The reason angular structure is generally ignored is that the ability to accurately constrain the angular structure of the gravitational field is nearly unique to gravitational lensing. Since we have not emphasized the ability of lenses to measure angular structure and other methods cannot do so very accurately, there has been little theoretical work on the angular structure of galaxies with dark matter. Both theoretical studies of halos and modelers of gravitational lenses need to pay more attention to the angular structure of the gravitational potential.

We start by analyzing a simple two-image lens using our non-parametric model of the monopole (Eqn. B.67) in an external shear (Eqn. B.51). The two images are located at $\boldsymbol{\theta}_A = \theta_A(\cos\chi_A, \sin\chi_A)$, and $\boldsymbol{\theta}_B = \theta_B(\cos\chi_B, \sin\chi_B)$ as illustrated in Fig. B.20. To illustrate the similarities and differences between shear and convergence, we will also include a global convergence κ_0 in the model. This corresponds to adding a term to the lens potential of the form $(1/2)\kappa_0\theta^2$. The model now has five parameters – two shear components, the mass and surface density of the monopole model and the additional global convergence. We have only two astrometric constraints, and so can solve for only two of the five parameters. Since the enclosed mass is always an interesting parameter, we can only solve for one of the two shear components. In general, we will find that the amplitude of γ_c depends on the amplitude of γ_s. There is, however, a special choice of the shear axis, $\chi_\gamma = (\chi_A + \chi_B)/2 + \pi/4$, such that the shear parameters become independent of each other. This allows us to determine the "invariant" shear associated with the images,

$$\gamma_1 = \frac{(1-\kappa_0 - \langle\kappa\rangle_{AB})\,(\theta_A^2 - \theta_B^2)\sin(\chi_A - \chi_B)}{\Delta\theta^2} \tag{B.78}$$

where $\Delta\theta = |\boldsymbol{\theta}_A - \boldsymbol{\theta}_B|$ is the image separation. The monopole mass and the other shear component are degenerate,

$$b_B^2 + \gamma_2\theta_A\theta_B = \frac{(1-\kappa_0)\left[\Delta\theta^2(\theta_A^2+\theta_B^2) - (\theta_A^2-\theta_B^2)^2\right] - \langle\kappa\rangle_{AB}\left(\theta_A^2-\theta_B^2\right)\left(\Delta\theta^2-\theta_A^2+\theta_B^2\right)}{2\Delta\theta^2}. \tag{B.79}$$

Several points are worth noting. First, the amplitude of the invariant shear γ_1 has the same degeneracy with the (local) surface density between the images $\langle\kappa\rangle_{AB}$ as it does with a global convergence κ_0. More centrally concentrated mass distributions with lower $\langle\kappa\rangle_{AB}$ require higher external shears to fit the same data. Second, the other component γ_2 introduces an uncertainty into the enclosed mass, with a series of somewhat messy trade offs between b_B^2, γ_2, $\langle\kappa\rangle_{AB}$ and κ_0. As a practical matter, the shear does not lead to an astronomically significant uncertainty in the mass, since $\gamma_2 \lesssim 0.1$ in all but the most extreme situations.

The external shear is only one component of the quadrupole. There is also an internal shear due to the mass interior to the images (Eqn. B.52). The internal and external shears differ in their "handedness". For the same angular deflection ($d\Psi/d\chi$) they have opposite signs for the radial deflection ($d\Psi/d\theta$). The solution for two images is much the same as for an external shear. There is an invariant shear component, whose amplitude scales with $1-\kappa_0-\langle\kappa\rangle_{AB}$ but whose orientation differs from that of the external shear solution. The monopole mass b_B^2 is degenerate with the γ_2 shear component and the κ_0 and $\langle\kappa\rangle_{AB}$ surface densities. The actual expressions are too complex to be illuminating.

Fig. B.23 illustrates how the invariant shears combine to determine the overall structure of the quadrupole for the lens PG1115+080. For each image pair there is a line of permitted shears because of the degeneracy between the enclosed mass and the second shear component. The invariant shear component is the shear at the point where the line passes closest to the origin. If the quadrupole model is correct, the lines for all the image pairs will cross at a point, while if it is incorrect they will not. PG1115+080 is clearly going to be well modeled if the quadrupole is dominated by an external shear and poorly modeled if it is dominated by an internal shear. This provides a simple geometric argument for why full models of PG1115+080 are always dominated by an external shear (e.g. Impey et al. 1998). A failure of the curves to cross in both cases is primarily evidence for a mixture of external and internal quadrupoles or the presence of other multipoles rather than for a problem in the monopole mass distribution. In Fig. B.23 we used an SIS for the monopole. For a point mass monopole, the figure looks almost the same provided we expand the scale – the invariant shear scales as $1-\langle\kappa\rangle_{AB}$ so in going from a SIS with $1-\langle\kappa\rangle_{AB} \simeq 1/2$ to a point mass with $1-\langle\kappa\rangle_{AB} = 1$ the shear will double.

This scaling of the quadrupole with the surface density of the monopole provides an as yet unused approach to studying the monopole. Since the mass

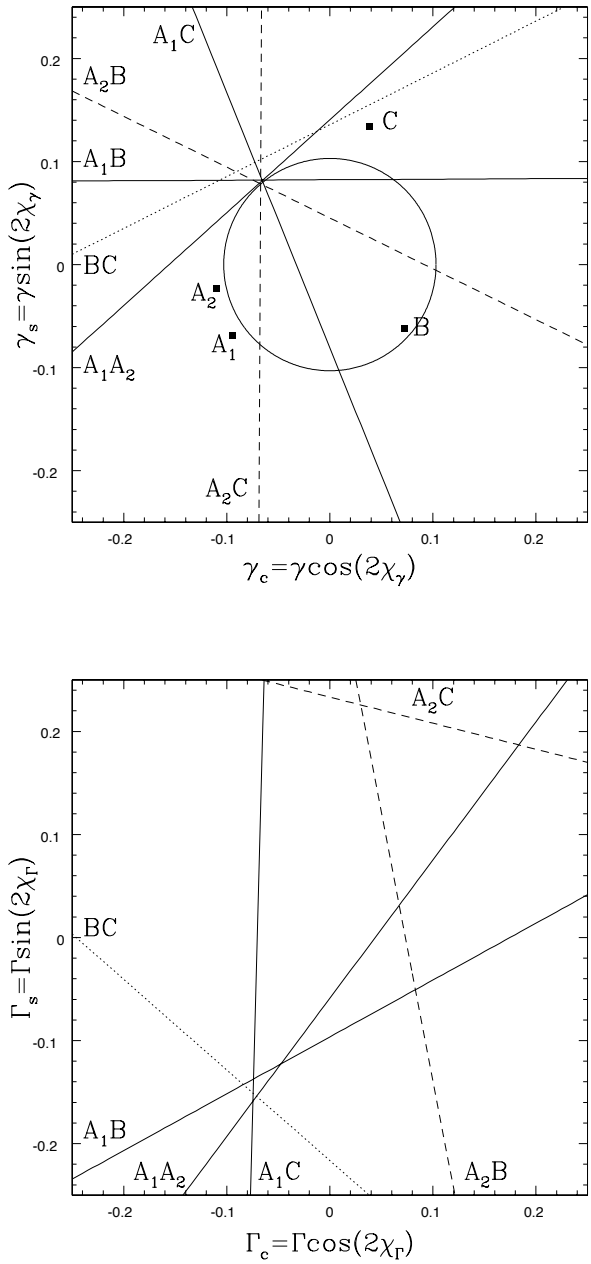


Fig. B.23. The invariant shears for the lens PG1115+080 modeled using either an external (top) or an internal (bottom) quadrupole and an SIS monopole. Each possible image pair among the A_1, A_2, B and C images, constrains the quadrupole to lie on the labeled line. The amplitude and orientation of each invariant shear is given by the point where the corresponding line passes closest to the origin. Models of PG1115+080 show that the quadrupole is dominated by external (tidal) shear. Here we see that for the external quadrupole (top), the lines nearly cross at a point, so the data are consistent with an almost pure external shear. For an internal quadrupole (bottom), the A_2B and A_2C image pairs require shear parameters completely inconsistent with the other images.

enclosed by the Einstein radius is nearly constant, the more centrally concentrated constant mass-to-light (M/L) ratio models must have lower surface mass densities near the images than the SIE model. As a result, they will require quadrupole amplitudes that are nearly twice those of models like the SIS with nearly flat rotation curves. Since the typical SIE model of a lens has an ellipticity that is comparable to the typical ellipticities of the visible galaxies, the more centrally concentrated monopole of a constant M/L model requires an ellipticity much larger than the observed ellipticity of the lens galaxy. The need to include an external tidal shear to represent the environment allows these models to produce acceptable fits, but the amplitudes of the required external shears are inconsistent with expectations from weak lensing (Part 3).

B.4.6 Model Fitting and the Mass Distribution of Lenses

Having outlined (in perhaps excruciating detail) how lenses constrain the mass distribution, we turn to the problem of actually fitting data. These days the simplest approach for a casual user is simply to down load a modeling package, in particular the lensmodel package (Keeton 2001a) at http://cfa-www.harvard.edu/castles/, read the manual, try some experiments, and then apply it intelligently (i.e. read the previous sections about what you can extract and what you cannot!). Please publish results with a complete description of the models and the constraints using standard astronomical nomenclature.

In most cases we are interested in the problem of fitting the positions $\boldsymbol{\theta}_i$ of $i = 1 \cdots n$ images where the image positions have been measured with accuracy σ_i. We may also know the positions and properties of one or more lens galaxies. Time delay ratios also constrain lens models but sufficiently accurate ratios are presently available for only one lens (B1608+656, Fassnacht et al. 2002), fitting them is already included in most packages, and they add no new conceptual difficulties. Flux ratios constrain the lens model, but we are so uncertain of their systematic uncertainties due to extinction in the ISM of the lens galaxy, microlensing (Part 4) and the effects of substructure (see §B.8) that we can never impose them with the accuracy needed to add a significant constraint on the model.

The basic issue with lens modeling is whether or not to invert the lens equations ("source plane" or "image plane" modeling). The lens equation supplies the source position

$$\boldsymbol{\beta}_i = \boldsymbol{\theta}_i - \boldsymbol{\alpha}(\boldsymbol{\theta}_i, \boldsymbol{p}) \tag{B.80}$$

predicted by the observed image positions $\boldsymbol{\theta}_i$ and the current model parameters $\boldsymbol{p}$. Particularly for parametric models it is easy to project the images on to the source plane and then minimize the difference between the projected

source positions. This can be done with a χ^2 fit statistic of the form

$$\chi^2_{src} = \sum_i \left(\frac{\boldsymbol{\beta} - \boldsymbol{\beta}_i}{\sigma_i} \right)^2 \tag{B.81}$$

where we treat the source position $\boldsymbol{\beta}$ as a model parameter. The astrometric uncertainties σ_i are typically a few milli-arcseconds. Moreover, where VLBI observations give significantly smaller uncertainties, they should be increased to approximately $0\farcs001$–$0\farcs005$ because low mass substructures in the lens galaxy can produce systematic errors on this order (see §B.8). *You can impose astrometric constraints to no greater accuracy than the largest deflection scales produced by lens components you are not including in your models.* The advantage of χ^2_{src} is that it is fast and has excellent convergence properties. The disadvantages are that it is wrong, cannot be used to compute parameter uncertainties, and may lead to a model producing additional images that are not actually observed.

The reason it is wrong and cannot be used to compute parameter errors is that the uncertainty σ_i in the image positions does not have any meaning on the source plane. This is easily understood if we Taylor expand the lens equation near the projected source point $\boldsymbol{\beta}_i$ corresponding to an image

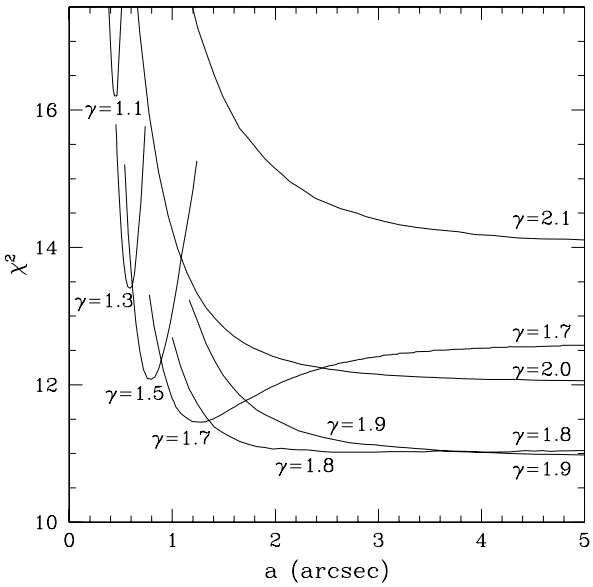
$$\boldsymbol{\beta} - \boldsymbol{\beta}_i = M_i^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_i) \tag{B.82}$$

where M_i^{-1} is the inverse magnification tensor at the observed location of the image. In the frame where the tensor is diagonal, we have that $\Delta\beta_\pm = \lambda_\pm \Delta\theta_\pm$ so a positional error $\Delta\beta_\pm$ on the source plane corresponds to a positional error $\lambda_\pm^{-1}\Delta\beta_\pm$ on the image plane. Since the observed lensed images are almost always magnified (usually $\lambda_+ = 1 + \kappa + \gamma \sim 1$ and $0.5 > |\lambda_- = 1 + \kappa - \gamma| < 0.05$) there is always one direction in which small errors on the source plane are significantly magnified when projected back onto the image plane. Hence, if you find solutions with $\chi^2_{src} \sim N_{dof}$ where N_{dof} is the number of degrees of freedom, you will have source plane uncertainties $\Delta\beta \lesssim \sigma_i$. However, the actual errors on the image plane are $\mu = |M|$ larger, so the χ^2 on the image plane is $\sim \mu^2 N_{dof}$ and you in fact have a terrible fit.

If you assume that in any interesting model you are close to having a good solution, then this Taylor expansion provides a means of using the easily computed source plane positions to still get a quantitatively accurate fitting statistic,

$$\chi^2_{int} = \sum_i \frac{(\boldsymbol{\beta} - \boldsymbol{\beta}_i) \cdot M_i^2 \cdot (\boldsymbol{\beta} - \boldsymbol{\beta}_i)}{\sigma_i^2}, \tag{B.83}$$

in which the magnification tensor M_i is used to correct the error in the source position to an error in the image position. This procedure will be approximately correct provided the observed and model image positions are close enough for the Taylor expansion to be valid. Finally, there is the exact statistic where for the model source position $\boldsymbol{\beta}$ you numerically solve the



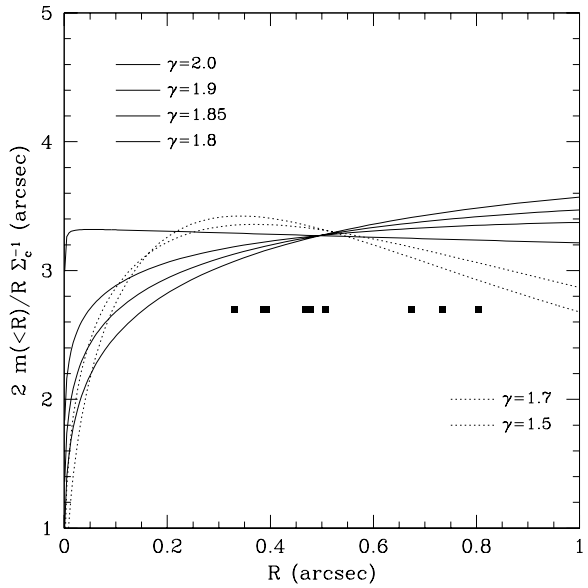


Fig. B.24. (Top) Goodness of fit χ^2 for cuspy models of B1933+503 as a function of the inner density exponent γ ($\rho \propto r^{-\gamma}$) and the profile break radius a. Models with cusps significantly shallower or steeper than isothermal are ruled out, and acceptable models near isothermal must have break radii outside the region with the lensed images.

Fig. B.25. (Bottom) The monopole deflections of the B1933+503 models for the range of permitted cusp exponents γ. The points show the radii of the lensed images, and the models only constrain the shape of the monopole in this region. The monopole deflection is closely related to the square of the rotation curve. Note the similarity to Fig. B.21.

lens equation to find the exact image positions $\boldsymbol{\theta}_i(\boldsymbol{\beta})$ and then compute the goodness of fit on the image plane

$$\chi^2_{img} = \sum_i \left(\frac{\boldsymbol{\theta}_i(\boldsymbol{\beta}) - \boldsymbol{\theta}_i}{\sigma_i} \right)^2 . \tag{B.84}$$

This will be exact even if the Taylor expansion of χ^2_{int} is breaking down, and if you find all solutions to the lens equations you can verify that the model predicts no additional visible images. Unfortunately, using the exact χ^2_{img} is also a much slower numerical procedure.

As we discussed earlier, even though lens models provide the most accurate mass normalizations in astronomy, they can constrain the mass distribution only if the source is more complex than a single compact component. Here we only show examples where there are multiple point-like components, deferring discussions of models with extended source structure to §B.10. The most spectacular example of a multi-component source is B1933+503 (Sykes et al. 1998, see Fig. B.7) where a source consisting of a radio core and two radio lobes has 10 lensed images because the core and one lobe are quadruply imaged and the other lobe is doubly imaged. Since we have many images spread over roughly a factor of two in radius, this lens should constrain the radial mass distribution just as in our discussion for §B.4.3. Muñoz et al. (2001, also see Cohn et al. 2001 for softened power law models) fitted this system with cuspy models (Eqn. 55 with $\alpha = 2$ and $m = 4$), varying the inner density slope $n = \gamma$ ($\rho \propto r^{-n}$) and the break radius a. Fig. B.24 shows the resulting χ^2 as a function of the parameters and Fig. B.25 illustrates the range of the acceptable monopole mass distributions – both are very similar to Fig. B.21. The best fit is for $\gamma = 1.85$ with an allowed range of $1.6 < \gamma < 2.0$ that completely excludes the shallow $\gamma = 1$ cusps of the Hernquist and NFW profiles and is marginally consistent with the $\gamma = 2$ cusp of the SIS model. A second example, which illustrates how the distribution of mass well outside the region with images has little effect on the models, are the Winn et al. (2003) models of the three-image lens PMNJ1632–0033 shown in Fig. B.26. In these models the outer slope η, with $\rho \propto r^{-\eta}$ asymptotically, of the density was also explored but has little effect on the results. Unless the break radius of the profile is interior to the B image, the mass profile is required to be close to isothermal $1.89 < \beta < 1.93$.

Unfortunately, systems like B1933+503 and PMNJ1632–0033 are a small minority of lens systems. For most lenses, obtaining information on the radial density profile requires some other information such as a dynamical measurement (§B.4.9), a time delay measurement (§B.5) or a lensed extended component of the source (§B.10). Even for these systems, it is important to remember that the actual constraints on the density structure really only apply over the range of radii spanned by the lensed images – the mass interior to the images is constrained but its distribution is not, while the mass exterior to the images is completely unconstrained. This is not strictly true

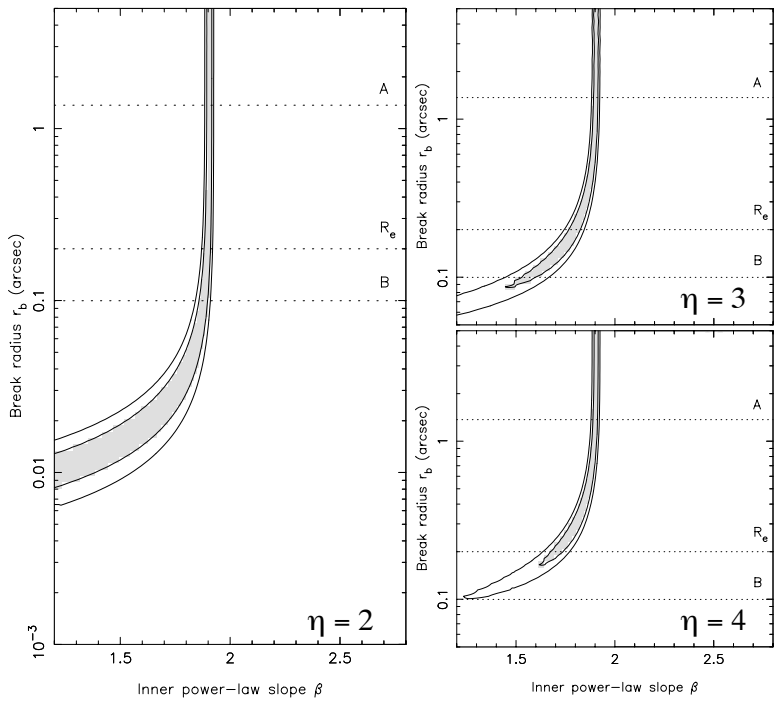


Fig. B.26. Allowed parameters for cuspy models of PMNJ1632–0033 assuming that image C is a true third image. Each panel shows the constraints on the inner density cusp β ($\rho \propto r^{-\beta}$) and the break radius r_b for three different asymptotic density slopes $\rho \propto r^{-\eta}$. A Hernquist model has $\beta = 1$ and $\eta = 4$, an NFW model has $\beta = 1$ and $\eta = 3$, and a pseudo-Jaffe model has $\beta = 2$ and $\eta = 4$. Unless the break radius is place interior to the B image, it is restricted to be close to isothermal ($\beta = 2$).

when we include the angular structure of the gravitational field and the mass distribution is quasi-ellipsoidal.

It is also important to keep some problems with parametric models in mind. First, models that lack the degrees of freedom needed to describe the actual mass distribution can be seriously in error. Second, models with too many degrees of freedom can be nonsense. We can illustrate these two limiting problems with the sad history of Q0957+561 for the first problem and attempts to explain anomalous flux ratios (see §B.8) with complex angular structures in the density distribution for the dark matter for the second.

Q0957+561, the first lens discovered (Walsh, Carswell & Weymann 1979) and the first lens with a well measured time delay (see §B.5, Schild & Thomson 1970, Kundić et al. 1997 and references therein), is an ideal lens for demonstrating the trouble you can get into using parametric models with-

out careful thought. The lens consists of a cluster and its brightest cluster galaxy with two lensed images of a radio source bracketing the galaxy. VLBI observations (e.g. Garrett et al. 1994) resolve the two images into thin, multi-component jets with very accurately measured positions (uncertainties as small as 0.1 mas, corresponding to deflections produced by a mass scale $\sim 10^{-8}$ of the primary lens!). Models developed along two lines. One line focused on models in which the cluster was represented as an external shear (e.g. Grogin & Narayan 1996, Chartas et al. 1998, Barkana et al. 1999, Chae 1999) while the other explored more complex models for the cluster (see Kochanek 1991b, Bernstein, Tyson & Kochanek 1993, Bernstein & Fischer 1999) and argued that external shear models had too few parameters to represent the mass distribution given the accuracy of the constraints. The latter view was born out by the morphology of the lensed host galaxy (Keeton et al. 2000) and direct X-ray observations of the cluster (Chartas et al. 2002) which showed that the lens galaxy was within about one Einstein radius of the cluster center where a tidal shear approximation fails catastrophically. The origin of the problem is that as a two-image lens, Q0957+561 is critically short of constraints unless the fine details of the VLBI jet structures are included in the models. Many studies imposed these constraints to the limit of the measurements while not including all possible terms in the potential which could produce a deflection on that scale (i.e. the precision should have been restricted to milli-arcseconds rather than micro-arcseconds). Models would adjust the positions and masses of the cluster and the lens galaxy in order to reproduce the small scale astrometric details of the VLBI jets without including less massive components of the mass distribution (e.g. the ellipticity gradient and isophote twist of the lens galaxy, Keeton et al. 2000) that also affect the VLBI jet structure on these angular scales. Lens models must contain all reasonable structures producing deflections comparable to the scale of the measurement errors.

We are in the middle of an experiment exploring the second problem – if you include small scale structures but lack the constraints needed to measure them, their masses easily become unreasonable unless constrained by common sense, physical priors or additional data. Lately this has become an issue in studies (Evans & Witt 2003, Kochanek & Dalal 2004, Quadri, Moller & Natarajan 2003, Kawano et al. 2004) of whether the flux ratio anomalies in gravitational lenses could be due to complex angular structure in the lens galaxy rather than CDM substructure or satellites in the lens galaxy (see §B.8). The problem, as we discuss in the next section on non-parametric models (§B.4.7), is that lens modeling with large numbers of parameters is closely related to solving linear equations with more variables than constraints – as the matrix inversion necessary to finding a solution becomes singular, the parameters of the mass distribution show wild, large amplitude fluctuations even as the fit to the constraints becomes perfect. Thus, a model including enough unconstrained parameters is guaranteed to "solve" the anomalous

flux ratio problem even if it should not. For example, Evans & Witt (2003) could match the flux ratios of Q2237+0305 even though for this lens we know from the time variability of the flux ratios that the flux ratio anomalies are created by microlensing rather than complex angular structures in the lens model (see Part 4).

If only the four compact images are modeled, then the flux ratio anomalies can be greatly reduced or eliminated in almost all lenses at the price of introducing deviations from an ellipsoidal density distribution far larger than expected (see §B.4.4). In some cases, however, you can test these solutions because the lens has extra constraints beyond the four compact images. We illustrate this in Fig. B.27. By adding large amplitude $\cos 3\theta$ and $\cos 4\theta$ perturbations to the surface density model for B1933+503, Kochanek & Dalal (2004) could reproduce the observed image flux ratios if they fit only the four compact sources. However, after adding the constraints from the other lensed components, the solution is driven back to being nearly ellipsoidal and the flux ratios cannot be fit. In every case, Kochanek & Dalal (2004) found that the extra constraints drove the solution back toward an ellipsoidal density distribution. In short, a sufficiently complex model can fit underconstrained data, but that does not mean it makes any sense to do so.

B.4.7 Non-Parametric Models

The basic idea behind non-parametric mass models is that the effective lens potential and the deflection equations are linear "functions" of the surface density. The surface density can be decomposed into multipoles (Kochanek 1991a, Trotter, Winn & Hewitt 2000, Evans & Witt 2003), pixels (see Saha & Williams 1997, 2004, Williams & Saha 2000), or any other form in which the surface density is represented as a linear combination of density functionals multiplied by unknown coefficients $\boldsymbol{\kappa}$. In any such model, the lens equation for image i takes the form

$$\boldsymbol{\beta} = \boldsymbol{\theta}_i - A_i \boldsymbol{\kappa} \tag{B.85}$$

where A_i is the matrix that gives the deflection at the position of image i in terms of the coefficients of the surface density decomposition $\boldsymbol{\kappa}$. For a lens with $i = 1 \cdots n$ images of the same source, such a system can be solved exactly if there are enough degrees of freedom in the description of the surface density. For simplicity, consider a two-image lens so that we can eliminate the source position by hand, leaving the system of equations

$$\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 = (A_1 - A_2)\boldsymbol{\kappa}, \tag{B.86}$$

which is easily solved by simply taking the inverse of the matrix $A_1 - A_2$ to find that

$$\boldsymbol{\kappa} = (A_1 - A_2)^{-1} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) . \tag{B.87}$$

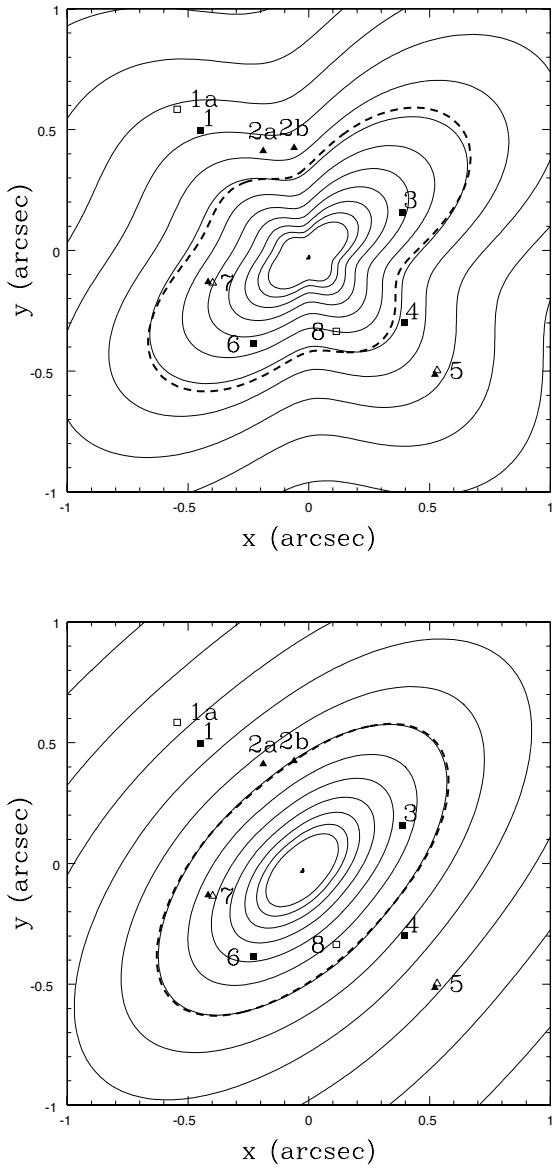


Fig. B.27. Surface density contours for models of B1933+503 including misaligned a_3 and a_4 multipoles (thin lines). The model in the top panel is constrained only by the 4 compact images (images 1, 3, 4 and 6, filled squares). The model in the bottom panel is also constrained by the other images in the lens (the two-image system 1a/8, open squares; the four-image system 2a/2b/5/7 filled triangles; and the two-image system comprising parts of 5/7, open triangles) The tangential critical line of the model (heavy solid curve) must pass between the merging images 2a/2b, but fails to do so in the first model (top panel).

Sadly, life is not that simple because as soon as the density decomposition has more degrees of freedom than there are constraints, the inverse $(A_1 - A_2)^{-1}$ of the deflection operators is singular.

The solution to this problem is to instead consider the problem as a more general minimization problem with a χ^2 statistic for the constraints and some form of regularization to restrict the results to plausible surface densities. One possibility is linear regularization, in which you minimize the function

$$F = \chi^2 + \lambda \boldsymbol{\kappa} \cdot H \cdot \boldsymbol{\kappa} \tag{B.88}$$

where the χ^2 measures the goodness of fit to the lens constraints, H is a weight matrix and λ is a Lagrange multiplier. The Lagrange multiplier controls the relative importance given to fitting the lens constraints (minimizing the χ^2) versus producing a smooth density distribution (minimizing $\boldsymbol{\kappa} \cdot H \cdot \boldsymbol{\kappa}$). The simplest smoothing function is to minimize the variance of the surface density ($H = I$, the identity matrix), or, equivalently, ignore H and use the singular value decomposition for inverting a singular matrix. By using more complicated matrices you can minimize derivatives of the density (gradients, curvature etc.). Solutions are found by adjusting the multiplier λ until the goodness of fit satisfies $\chi^2 \simeq N_{dof}$ where N_{dof} is the number of degrees of freedom. Another solution is to use linear programming methods to impose constraints such as positive surface densities, negative density gradients from the lens center or density symmetries (Saha & Williams 1997, 2004, Williams & Saha 2000). Time delays, which are also linear functions of the surface density, are easily included. Flux ratios are more challenging because magnifications are quadratic rather than linear functions of the surface density except for the special case of the generalized singular isothermal models where $\Psi = b\theta F(\chi)$ (Eqn. B.42, Witt, Mao & Keeton 2000, Kochanek et al. 2001, Evans & Witt 2001). The best developed, publicly available non-parametric models are those by Saha & Williams (2004). These are available at http://ankh-morpork.maths.qmc.ac.uk/∼saha/astron/lens/.

Personally, I am not a fan of the non-parametric models, because almost all the additional degrees of freedom they include are irrelevant to the problem. As I have tried to outline in the preceding sections, there is no real ambiguity about the aspects of gravitational potentials either constrained or unconstrained by lens models. Provided the parametric models capture these degrees of freedom and you do not get carried away with the precision of the fits, you can ignore deviations of the $\cos(16\chi)$ term of the surface density from that expected for an ellipsoidal model. Similarly for the monopole profile, the distribution of mass interior and exterior to the images is irrelevant and for the most part only the mean surface density between the images has any physical effect. Nothing is gained by allowing arbitrary, fine-grained distributions.

There are also specific physical and mathematical problems with non-parametric models just as there are for parametric models. First, the trick of

linearization only works if the lens equations are solved on the source plane. As we discussed when we introduced model fitting (§B.4.6), this makes it impossible to properly compute error bars on any parameters. The equations become non-linear if they include either the magnification tensor (Eqn. B.83) or use the true image plane fit statistic (Eqn. B.84), and this greatly reduces the attractiveness of these methods. Second, in many cases the non-parametric models are not constrained to avoid creating extra images not seen in the observations – the models reproduce the observed images exactly, but come with no guarantee that they are not producing 3 other images somewhere else. Third, it is very difficult to guarantee that the resulting models are physical. For example, consider a simple spherical lens constrained to have positive surface density. For the implied three-dimensional density to also be positive definite, the surface density must decline monotonically from the center of the lens. This constraint is usually applied by the Saha & Williams (2004) method. For the distribution function of the stars making up the galaxy to be positive definite, the three dimensional density must also decline monotonically – this implies a constraint on the second derivative of the surface density which is not imposed by any of these methods. For the distribution to be dynamically stable it must satisfy a criterion on the derivative of the distribution function with respect to the orbital energy, and this implies a criterion on the third derivative of the surface density which is also not imposed (see Binney & Tremaine 1987). Worse yet, for a non-spherical system we cannot even write down the constraints on the surface density required for the model to correspond to a stable galaxy with a positive definite distribution function. In short, most non-parametric models will be unphysical – they overestimate the degrees of freedom in the mass distribution. The critique being made, parametric models have a role because they define the outer limits of what is possible by avoiding the strong physical priors implicit in parametric models of galaxies.

B.4.8 Statistical Constraints on Mass Distributions

Where individual lenses may fail to constrain the mass distribution, ensembles of lenses may succeed. There are two basic ideas behind statistical constraints on mass distributions. The first idea is that models of individual lenses should be weighted by the likelihood of the observed configuration given the model parameters. The second idea is that the statistical properties of lens samples should be homogeneous.

An example of weighting models by the likelihood is the limit on the slopes of central density cusps from the observed absence of central images. Rusin & Ma (2001) considered 6 CLASS (see §B.6) survey radio doubles and computed the probability $p_i(n)$ that lens i would have a detectable third image in the core of the lens assuming power law mass densities $\Sigma \propto R^{1-n}$ and including a model for the observational sensitivities and the magnification bias (see §B.6.6) of the survey. They were only interested in the range $n < 2$, because

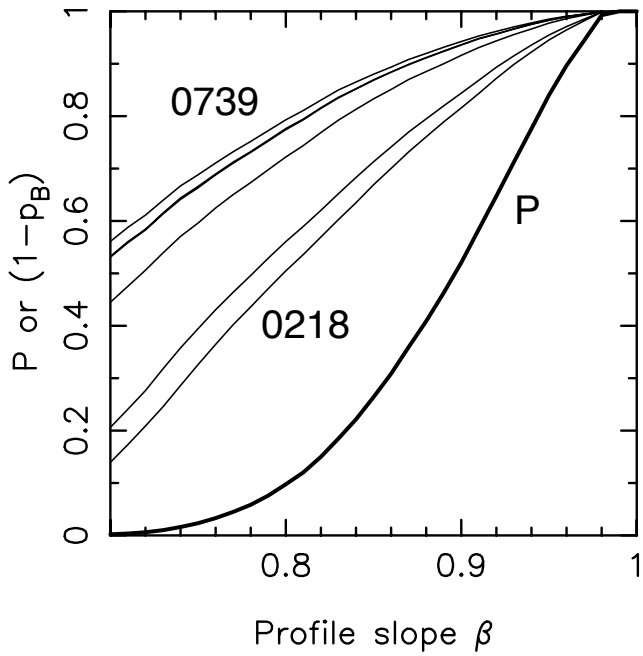
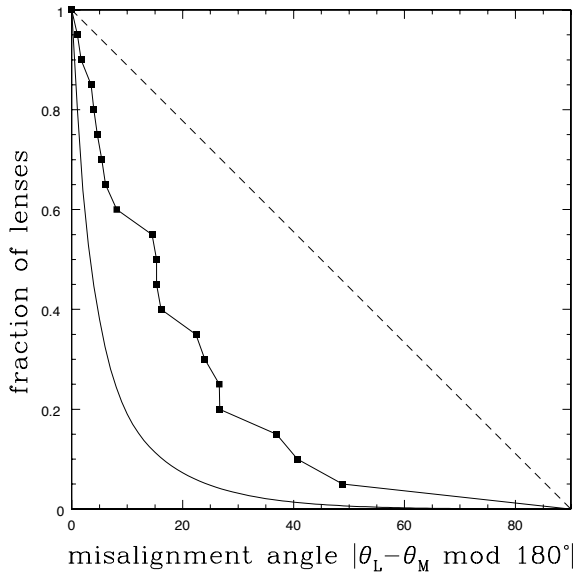


Fig. B.28. Limits on the central density exponent for power-law density profiles $\rho \propto r^{-n} = r^{-1-\beta}$ from the absence of detectable central images in a sample of 6 CLASS survey radio doubles (Rusin & Ma 2001). The lighter curves show the limits for the individual lenses with the weakest constraint from B0739+366 and the strongest from B0218+357, and the heavy solid curve shows the joint probability P.

as discussed in §B.3, density cusps with $n \geq 2$ never have central images. For most of the lenses they considered, it was possible to find models of the 6 lenses that lacked detectable central images over a broad range of density exponents. However, the shallower the cusp, the smaller the probability $p_i(n)$ of producing a lens without a visible central image. For any single lens, $p_i(n)$ varies too little to set a useful bound on the exponent, but the joint probability of the entire sample having no central images, $P = \Pi_i(1-p_i(n))$, leads to a strong (one-sided) limit that $n > 1.78$ at 95% confidence (see Fig. B.28). In practice, Keeton (2003a) demonstrated that the central stellar densities are sufficiently high to avoid the formation of visible central images in almost all lenses given the dynamic ranges of existing radio observations (i.e. stellar density distributions are sufficiently cuspy), and central black holes can also assist in suppressing the central image (Mao, Witt & Koopmans 2001). However, the basic idea behind the Rusin & Ma (2001) analysis is important and underutilized.

An example of requiring the lenses to be homogeneous is the estimate of the misalignment between the major axis of the luminous lens galaxy



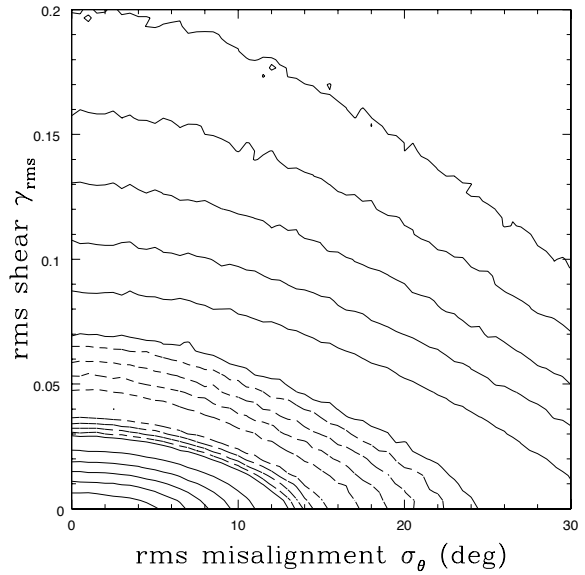


Fig. B.29. (Top) The integral distribution of misalignment angles $\Delta\chi_{LM}$ between the major axes of the lens galaxy and an ellipsoidal lens model (solid curve with points for each lens). If the two angles were completely uncorrelated, the distribution would follow the dashed line. If the two angles were perfectly correlated they would follow the solid curve because of the measurement uncertainties in the two angles.
Fig. B.30. (Bottom) Logarithmic contours of the probability for matching the distribution of misalignment angles as a function of the rms misalignment σ_θ between the mass and the light and the typical tidal shear γ_{rms}. Theoretically we expect tidal shears $\gamma_{rms} \simeq 0.06$. The solid contours are spaced by 0.5 dex and the dashed contours are spaced by 0.1 dex relative to the maximum likelihood contour. The differences between dashed contours are not statistically significant, while those between solid contours are statistically significant.

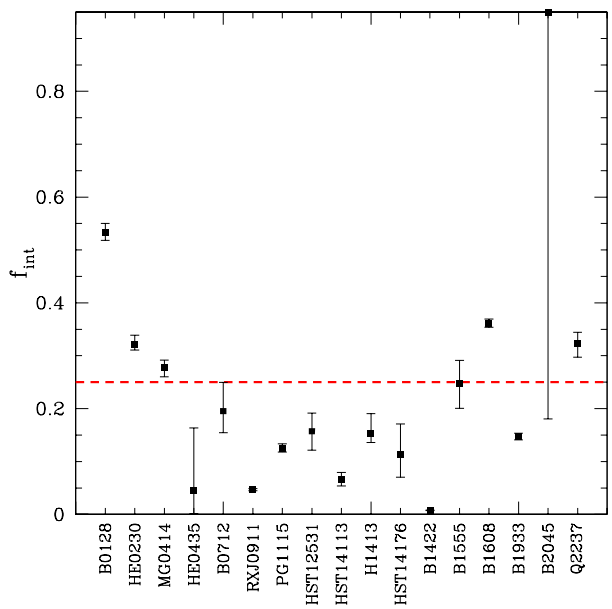


Fig. B.31. The internal shear fraction f_{int} for the four-image lenses. Each system was fitted by an SIS combined with an internal shear and an external shear and $f_{int} = |\Gamma|/(|\Gamma|+|\gamma|)$ is the fraction of the quadrupole amplitude due to the internal shear. An SIE has $f_{int} = 1/4$ (see Fig B.22). Most of the quads have $f_{int} \lesssim 1/4$ as expected for an SIE in an additional external (tidal) shear field. Objects with very low f_{int} (e.g. HE0435–1223, RXJ0911+0551, B1422+231) have nearby galaxies or clusters generating anomalously large external shears, while objects with anomalously high f_{int} (B1608+656, HE0230–2130, MG0414+0534) tend to have additional lens components like the second lens galaxy of B1608+656. For some systems either the imaging data (e.g. B0128+437) or the models (e.g. B2045+265) do not allow a clear qualitative explanation.

and the overall mass distribution by Kochanek (2002b). Fig. B.29 shows the misalignment angle $\Delta\chi_{LM} = |\chi_L - \chi_M|$ between the major axis χ_L of the lens galaxy and the major axis χ_M of an ellipsoidal mass model for the lens. The particular mass model is unimportant because any single component model of a four-image lens will give a nearly identical value for χ_M (e.g. Kochanek 1991a, Wambsganss & Paczynski 1994). The distribution of the misalignment angle $\Delta\chi_{LM}$ is not consistent with the mass and the light being either perfectly correlated or uncorrelated. This is not surprising, because a simple ellipsoidal model determines the position angle of the mean quadrupole moment near the Einstein ring, which is a combination of the quadrupole moment of the lens galaxy, the halo of the lens galaxy, and the local tidal shear (see §B.4.4). Even if the lens galaxy and the halo were perfectly aligned,

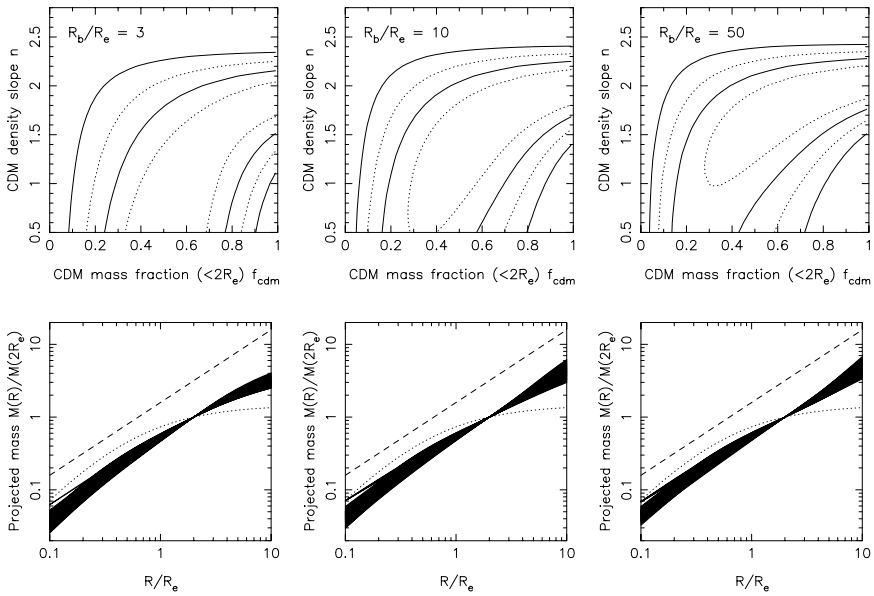


Fig. B.32. The structure of lens galaxies in self-similar models. The top row shows the permitted region for the slope of the inner dark matter cusp ($\rho \propto r^{-n}$) and the projected fraction of the mass f_{CDM} inside $2R_e$ composed of dark matter. The results are shown for three ratios R_b/R_e between the break radius R_b of the dark matter profile and the effective radius R_e of the luminous galaxy. The solid (dashed) contours show the 68% and 95% confidence levels for two (one) parameter. Note that the estimates of n and f_{CDM} depend little on the location of the break radius relative to the effective radius. The bottom row shows all the mass profiles lying within the (two parameter) 68% confidence region normalized to a fixed projected mass inside $2R_e$. For comparison we show the mass enclosed by a de Vaucouleurs model (dotted line) and an SIS (offset dashed line). While the allowed models exhibit a wide range of dark matter abundances, slopes and break radii, they all have roughly isothermal total mass profiles over the radial range spanned by the lensed images.

we would still find that the orientation of the mean quadrupole would differ from that of the light because of the effects of the tidal shears. We can model this by estimating the probability of reproducing the observed misalignment distribution in terms of the strength of the local tidal shear γ_{rms} and the dispersion in σ_χ in the angle between the major axis of the mass distribution and the light, as shown in Fig. B.30. The observed mismatch can either be produced by having a typical tidal shear of $\gamma_{rms} \simeq 0.05$ or by having a typical misalignment between mass and light of $\sigma_\chi \simeq 20°$. We know, however, that the typical tidal shear cannot be zero because it can be estimated from the statistics of galaxies (e.g. Keeton, Kochanek & Seljak 1997, Holder & Schechter 2003). Keeton et al. (1997) obtained $\gamma_{rms} \simeq 0.05$, in which case

mass must align with light and we obtain an upper limit of $\sigma_\chi \lesssim 10°$. Holder & Schechter (2003) argue for a much higher rms shear of $\gamma_{rms} = 0.15$ based on N-body simulations, which is too high to be consistent with the observed alignment of mass models and the luminous galaxy. One possible explanation (based on the results of White, Hernquist & Springel 2001) is that Holder & Schechter (2003) included parts of the lens galaxy's own halo in their estimate of the external shear. Alternatively, if lens galaxies are more compact than the SIE model used by Kochanek (2002b), then the lower surface density $\langle\kappa\rangle$ raises the required shear (since $\gamma \propto (1 - \langle\kappa\rangle)$, Eqn. B.78). However, mass distributions similar to constant mass-to-light ratio models of the lenses would be required, which would be inconsistent with shear estimates from simulations in which galaxy masses are dominated by extended dark matter halos.

The trade-off between central concentration and shear leads to the the interesting question of where the quadrupole structure of lenses originates. As we discussed in §B.4.4, we can break up the quadrupole of the mass distribution into the internal quadrupole due to the matter interior to the Einstein ring (Eqn. B.52) and the exterior quadrupole due to the matter outside the Einstein ring (Eqn. B.51). While the internal quadrupole is due only to the lens galaxy, the external quadrupole is a mixture of the quadrupole from the parts of the galaxy outside the Einstein ring (i.e. the dark matter halo) and the tidal shear from the environment. An important fact to remember is that for an isothermal ellipsoid, only $f_{int} = 25\%$ of the quadrupole is due to mass inside the Einstein ring (see Fig. B.22, §B.4.4)! Turner, Keeton & Kochanek (2004) explored this by fitting all the available four-image lenses with an SIS monopole combined with an internal and an external quadrupole. They then computed the fraction of the quadrupole f_{int} associated with the mass interior to the Einstein ring to find the distribution shown in Fig. B.31. Most four-image lenses seem to be dominated by the external quadrupole, with internal quadrupole fractions below the $f_{int} = 0.25$ fraction expected for an isothermal ellipsoid. Lenses clearly in environments with very large tidal shears (e.g. RXJ0911+0551 which is near massive cluster, Bade et al. 1997, Kneib et al. 2000, Morgan et al. 2001 or HE0435–1223 which is near a large galaxy, Wisotzki et al. 2002, see Fig. B.4) show much smaller internal shear fractions. B1608+656 (Myers et al. 1995, Fassnacht et al. 1999), which has two lens galaxies inside the Einstein ring, shows a significantly higher internal quadrupole fraction. Combined with the close correlation of mass model alignments with the luminous galaxies, this seems to argue for significant dark matter halos aligned with the luminous galaxy, but the final step of quantitatively assembling all the pieces has yet to be done.

Statistical analyses can also be used to estimate the radial density distribution from samples of lenses which individually cannot. The existence of the fundamental plane (see §B.9) strongly suggests that the structure of early-type galaxies is fairly homogeneous – in particular it is consistent

with galaxies having self-similar mass distributions in the sense that the halo structure can be scaled from the structure of the visible galaxy. As a particular example based on our theoretical expectations, Rusin, Kochanek & Keeton (2003) and Rusin & Kochanek (2004) modeled the visible galaxy with a Hernquist (Eqn. B.56) model scaled to match the observed effective radius of the lens galaxy, R_e, and then added a cuspy dark matter halo (Eqn. B.59 with a variable inner cusp η, $\alpha = 2$ and $m = 3$) where the inner density cusp ($\rho \propto r^{-\eta}$), the halo break radius r_b and the dark matter fraction f_{CDM} inside $2R_e$ were kept as variables. The assumption of self-similarity enters by keeping the ratio r_b/R_e constant, the dark matter fraction f_{CDM} constant, and then scaling the mass-to-light ratio of the stars $\Upsilon \propto L^x$ with the luminosity.[4] We recover the fundamental plane in this model when $x \simeq 0.25$. Putting all the pieces together, the projected mass inside radius R is

$$M(<R) = \Upsilon_* L_* \left(\frac{L(0)}{L_*}\right)^{1+x} \left[g(R/R_e) + g(2)\frac{f_{CDM}}{1-f_{CDM}} m_{CDM}(R/R_e)\right] \tag{B.89}$$

where Υ_* is the mass-to-light ratio of the stars in an L_* galaxy, $\log L(0) = \log L(z) - e(z)$ is the luminosity of the lens galaxy evolved to redshift zero (where we discuss estimates of the evolution rate $e(z)$ in §B.9), $g(x)$ is the fraction of the light inside dimensionless radius $x = R/R_e$ ($g(1) = 1/2$) and $m_{CDM}(x)$ is the dimensionless dark matter mass inside radius x with $m_{CDM}(2) = 1$ so that the CDM mass fraction inside $x = 2$ is f_{CDM}.

As we discussed earlier in §B.4.6, few lenses have sufficient constraints to estimate all the parameters in such a complex model. However, the assumption of self-similarity allows the average profile to be constrained statistically (Rusin et al. 2003, 2004). Suppose we saw lensed images generated by the same galaxy at a range of different source and lens redshifts. Each observed lens only reliably measures an aperture mass $M_{ap}(R < R_{Ein}) = \pi \Sigma_c R_{Ein}^2$ where R_{Ein} is the Einstein radius. But the physical scale R_{Ein} varies with redshift, so the ensemble of the lenses traces out the overall mass profile. Clearly we do not have ensembles of lenses generated by identical galaxies, but the assumption of self-similarity allows us to use the same idea for lenses with a range of luminosities and scale lengths. For 22 lenses with redshifts and accurate photometry we compared the measured aperture masses to the predicted aperture masses (the procedure for two-image lenses is a little more complicated, see Rusin et al. 2003) to estimate all the model parameters. Fig. B.32 shows the results for the parameters associated with the dark matter halo. In the limit that $f_{CDM} \to 1$ we find that the mass distribution is consistent with a simple SIS model (the limit $f_{CDM} \to 1$ and $n \to 2$) almost independent of the break radius location. There is a slight trend with break radius because as the break to the steep $\rho \propto r^{-3}$ outer profile gets closer

[4] They could also have allowed the CDM fraction to vary as $f_{CDM} \propto L^y$, but these led to degenerate models where only the combination $x + y$ was constrained.

to the region with the lensed images the inner cusp can be shallower while keeping the overall profile over the region with images close to isothermal. As we reduce f_{CDM} and add mass to the stars, the inner cusp becomes shallower, such that for a NFW ($n = 1$) cusp the dark matter fraction inside $2R_e$ is $\sim 40\%$. It is interesting to note, however, that the total mass distribution (light + dark) changes little over the full range of allowed parameters (bottom panels of Fig. B.32) – lensing constrains the global mass distribution not how it is divided into luminous and dark subcomponents. Note the resemblance of the statistical results to the results for detailed models of B1933+503 in Fig. B.25.

B.4.9 Stellar Dynamics and Lensing

Stellar dynamical analyses of gravitational lenses have reached the level of studies of local galaxies approximately 15–20 years ago. The analyses are based on the spherical Jeans equations (see Binney & Tremaine 1987) with simple models of the orbital anisotropy and generally ignore both deviations from sphericity and higher order moments of the velocity distributions. The spherical Jeans equation

$$\frac{1}{\nu}\frac{d\nu\sigma_r^2}{dr} + \frac{2\beta(r)}{r}\sigma_r^2 = -\frac{GM(r)}{r^2} \tag{B.90}$$

relates the radial velocity dispersion $\sigma_r = \langle v_r^2\rangle^{1/2}$, the isotropy parameter $\beta(r) = 1 - \sigma_\theta^2/\sigma_r^2$ characterizing the ratio of the tangential dispersion to the radial dispersion, the luminosity density of the stars $\nu(r)$ and the mass distribution $M(r)$. A well known result from dynamics is that you cannot infer the mass distribution $M(r)$ without constraining the isotropy $\beta(r)$ (e.g. Binney & Mamon 1982). Models with $\beta = 0$ are called isotropic models (i.e. $\sigma_r = \sigma_\theta$), while models with $\beta \to 1$ are dominated by radial orbits ($\sigma_\theta \to 0$) and models with $\beta \to -\infty$ are dominated by tangential orbits ($\sigma_r \to 0$) . These 3D components of the velocity dispersion must then be projected to measure the line-of-sight velocity dispersion $\langle v_{los}^2\rangle^{1/2}$,

$$\Sigma(R)\langle v_{los}^2\rangle(R) = 2\int_0^\infty dz\nu\left[\frac{z^2}{r^2}\sigma_r^2 + \frac{R^2}{r^2}\sigma_\theta^2\right] = 2\int_0^\infty dz\nu\sigma_r^2\left[1 - \frac{R^2}{r^2}\beta(r)\right] \tag{B.91}$$

where $\Sigma(R) = 2\int_0^\infty dz\nu(r)$ is the projected surface brightness and z is the coordinate along the line of sight.

Modern observations of local galaxies break the degeneracy between mass and isotropy by measuring higher order moments ($\langle v_{los}^n\rangle$) of the line-of-sight velocity distribution (LOSVD) because the shape of the LOSVD is affected by the isotropy of the orbits. Because the velocity dispersions are measured starting from a Gaussian fit to the LOSVD, the higher order moments are described by the amplitudes h_n of a decomposition of the LOSVD into Gauss-Hermite polynomials (e.g. van der Marel & Franx 1993). In general, the rms

velocity (i.e. combining dispersion and rotation) and higher order moment profiles of early-type galaxies are fairly self-similar, with nearly flat rms velocity profiles, modest values of $h_4 \simeq 0.01 \pm 0.03$ and slightly radial orbits $\langle\beta\rangle \simeq 0.1$–$0.2$ (e.g. Romanowsky & Kochanek 1999, Gerhard et al. 2001).

Stellar dynamics is used for two purposes in lensing studies. The first is to provide a mass normalization for lens models used in studies of lens statistics. We will discuss this in §B.6. The second is to use comparisons between a mass estimated from the geometry of a lens and the velocity dispersion of the lens galaxy to constrain the mass distribution (e.g. Romanowsky & Kochanek 1999, Trott & Webster 2002, Koopmans & Treu 2002, 2003, Treu & Koopmans 2002a, 2002b, Koopmans et al. 2003). It is important to understand that the systematic uncertainties in combining lensing and stellar dynamics to determine mass distributions are different from using either in isolation. For local galaxies we measure a velocity dispersion profile. The normalization of the profile sets the mass scale and the changes in the profile (and any higher order moments) with radius constrains the mass distribution. To lowest order, a simple scaling error in the velocity measurements will lead to errors in the mass scale rather than in the mass distribution. For lens galaxies, it is the comparison between the velocity dispersion and the mass determined by the geometry of the images that constrains the mass distribution. Thus, estimates of the mass distribution are directly affected by any calibration errors in the velocity dispersions.

We can understand the differences with a simple thought experiment. Suppose we have a mass distribution $M = M_0(R/R_0)^x$ in projection and we have mass estimates M_1 at R_1 and M_2 at R_2. Combining them we can solve for the exponent describing the mass distribution, $x = \ln(M_1/M_2)/\ln(R_1/R_2)$. In a dynamical observation the mass estimate is some sort of virial estimator $M \propto \sigma_v^2 R/G$ while in a lensing measurement it is a direct measurement of M. Standard velocity dispersion measurements start from the best fit Gaussian line width $\hat{\sigma}$ and then subtract an intrinsic line width σ_c due to the instrument and the intrinsic line width of the star in quadrature to estimate the portion of the line width due to the motions of the stars. Thus $\sigma_v^2 = f^2(\hat{\sigma}^2 - \sigma_c^2)$ where $f \simeq 1$ is a scale factor to account for deviations from spherical symmetry and non-Gaussian line of sight velocity distributions (LOSVDs). In a purely dynamical study, uncertainties in f and σ_c produce bigger fractional errors in the absolute mass scale M_0 than in the exponent x. For example, given measurements σ_1 and σ_2 at radii R_1 and R_2, the exponent, $x = 1 + \ln(\sigma_1^2/\sigma_2^2)/\ln(R_1/R_2)$, depends only on velocity dispersion ratios in which calibration errors tend to cancel. This is obvious for the scale factor f, which cancels exactly if it does not vary with radius. Since studies of lens dynamics use a comparison between a dynamical mass and a lensing mass to estimate the mass distribution, the results are more sensitive to calibration problems because these cancellations no longer occur. If we combine a velocity dispersion measurement σ_1 with a lensing mass measurement M_2 our

estimate of the exponent becomes $x = \ln(\sigma_1^2 R_1/GM_2)/\ln(R_1/R_2)$ and the uncertainties are linear in the scale factor f rather than canceling. An error analysis for the effects of σ_c is messier, but you again find that the sensitivity in the mixed lensing and dynamics constraint to errors in σ_c is greater than in a purely dynamical study.

Velocity dispersions have now been measured for 10 lenses (0047–2808 Koopmans & Treu 2003; CFRS03.1077 Treu & Koopmans 2004; Q0957+561 Falco et al. 1997, Tonry & Franx 1999; PG1115+080 Tonry 1998); HST14176+5226 Ohyama et al. 2002, Gebhardt et al. 2003, Treu & Koopmans 2004; HST15433+5352 Treu & Koopmans 2004; MG1549+3047 Lehár et al. 1996; B1608+656 Koopmans et al. 2003; MG2016+112 Koopmans & Treu 2002; Q2237+0305 Foltz et al. 1992). With the exception of Romanowsky & Kochanek (1999), who fitted for the distribution function of the lens, the analyses of the data have used the spherical Jeans equations with parameterized models for the isotropy $\beta(r)$. They include the uncertainties in σ_c about as well as any other dynamical study, although it is worth bearing in mind that this is tricky because we lack nearby stars with the appropriate metallicity and the problem of matching the spectral resolution for the galaxy and the template stars lacks direct checks of the success of the procedure. A useful rule of thumb to remember is that repeat measurements of velocity dispersions by different groups almost always show larger scatter than is consistent with the reported uncertainties. For example, the three velocity dispersion measurements for the lens HST14176+5226 (224 ± 15 km/s by Ohyama et al. 2002, 202 ± 9 km/s by Gebhardt et al. 2003, and 230 ± 14 km/s by Treu & Koopmans 2004) are mutually consistent only if the uncertainties are broadened by 30%.

In Fig. B.33 we summarize the dynamical constraints for 9 of these systems using the self-similar mass distribution from Rusin & Kochanek (2004, Eqn. B.89). This model is very similar to that used by Treu & Koopmans (2004). For most of the lenses, the region producing a good fit to the combined lensing and dynamical data overlaps the same region preferred by the Rusin & Kochanek (2004) self-similar models, shows the same general parameter degeneracy and is consistent with a simple SIS mass distribution ($f_{cdm} \to 1$ and $n = 2$). This is particularly true of 0047–2808, HST15433+5352, B1608+656, MG2016+112 and CFRS03.1077. Only Q2237+0305, where the lens is the bulge of a nearby spiral and we might not expect this mass model to be applicable, shows a very different trend (e.g. see the models of Trott & Webster 2002). PG1115+080 and to a lesser extent MG1549+3047 might have steeper than isothermal mass distributions (falling rotation curves) and the possibility of being consistent with a constant mass-to-light ratio model (Treu & Koopmans 2002a). HST14176+5226 and to a lesser extent HST15433+5352 could have shallower than isothermal mass distribution (rising rotation curves). Along the degeneracy direction for each lens we will find similar mass distributions with very different decompositions into luminous and dark matter, just as in Fig. B.32. The problem raised by this panorama

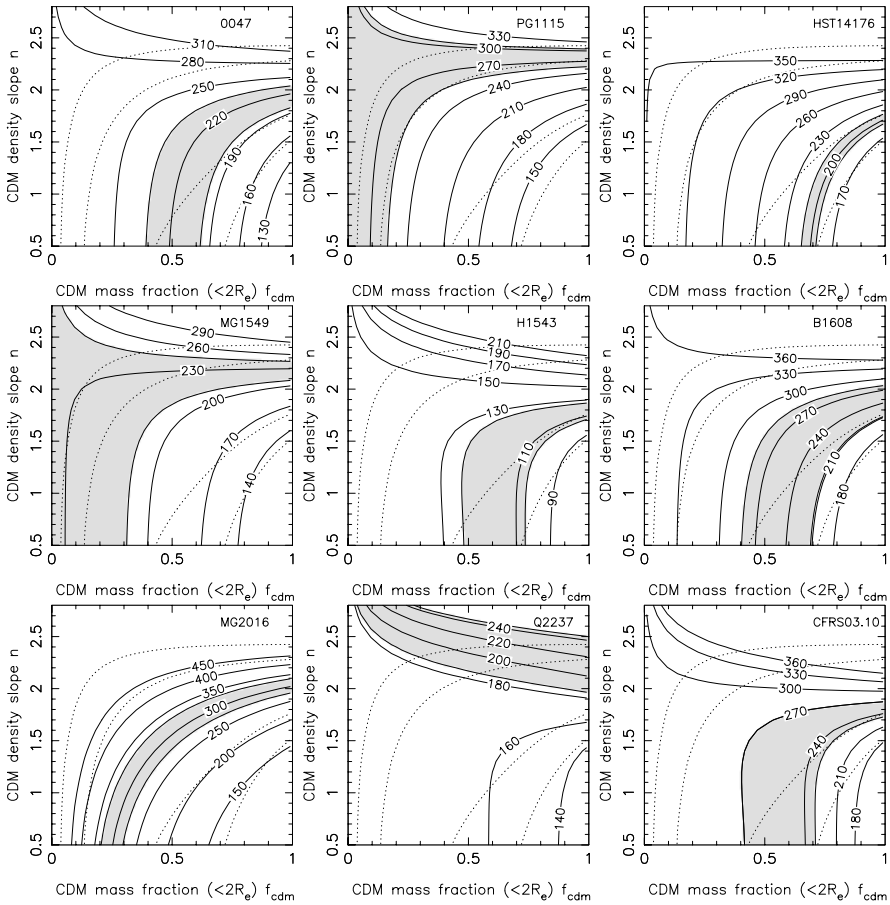


Fig. B.33. Constraints from lens velocity dispersion measurements on the self-similar mass distributions of Eqn. B.89 and Fig. B.32. The dotted contours show the 68% and 95% confidence limits from the self-similar models for $R_b/R_e = 50$. The shaded regions show the models allowed (68% confidence) by the formal velocity dispersion measurement errors, and the heavy solid lines show contours of the velocity dispersion in km/s. We used the low Gebhardt et al. (2003) velocity dispersion for HST14176+5226 because it has the smallest formal error. These models assumed isotropic orbits, thereby underestimating the full uncertainties in the stellar dynamical models.

is whether it shows that the halo structure is largely homogeneous with some measurement outliers, or that the structure of early-type is heterogeneous with important implications for understanding time delays (§B.5) and galaxy evolution (§B.9).

My own view tends toward the first interpretation – that the dynamical data supports the homogeneity of early-type galaxy structure. The permitted bands in Fig. B.33 show the 68% confidence regions given the formal measurement errors and the simple, spherical, isotropic Jeans equation models – this means that the true 68% confidence regions are significantly larger. We have already argued that the formal errors on dynamical measurements tend to be underestimates. For example, the need for HST14176+5226 to have a rising rotation curve would be considerably reduced if we used the higher velocity dispersion measurements from Ohyama et al. (2002) or Treu & Koopmans (2004) or if we broadened the uncertainties by the 30% needed to make the three estimates statistically consistent. Moreover, the existing analyses have also neglected the systematic uncertainties arising from the scaling factor f. There are two important issues that make $f \neq 1$. The first issue is that standard velocity dispersion measurements are the width of the best fit Gaussian model for the LOSVD, and this is not the same as the mean square velocity ($\langle v_{los}^2 \rangle^{1/2}$) appearing in the Jeans equations used to analyze the data unless the LOSVD is also a Gaussian. Stellar dynamics has adopted the dimensionless coefficients h_n of a Gauss-Hermite polynomial series to model the deviations of the LOSVD from Gaussian, and a typical early-type galaxy has $|h_4| \lesssim 0.03$ (e.g. Romanowsky & Kochanek 1999). This leads to a systematic difference between the measured dispersions and the mean square velocity of $\langle v_{los}^2 \rangle^{1/2} \simeq \sigma(1 + \sqrt{6}h_4)$ (e.g. van der Marel & Franx 1993)), so $|f - 1| \sim 7\%$ for $|h_4| \simeq 0.03$. Only the Romanowsky & Kochanek 1999) models of Q0957+561 and PG1115+080 have properly included this uncertainty. In fact, Romanowsky & Kochanek (1999) demonstrated that there were stellar distribution functions in which the mass distribution of PG1115+080 is both isothermal and agrees with the measured velocity dispersion. While it is debatable whether these models allowed too much freedom, it is certainly true that models using the Jeans equations and ignoring the LOSVD have too little freedom and will overestimate the constraints.

The second issue is that lens galaxies are not spheres. Unfortunately there are few simple analytic results for oblate or triaxial systems like early-type galaxies in which the ellipticity is largely due to anisotropies in the velocity dispersion tensor rather than rotation. For the system as a whole, the tensor virial theorem provides a simple global relationship between the major and minor axis velocity dispersions

$$\frac{\sigma_{major}}{\sigma_{minor}} \simeq 1 + \frac{1}{5}e^2 + \frac{9}{70}e^4 + \cdots \tag{B.92}$$

for an oblate ellipsoid of axis ratio q and eccentricity $e = (1 - q^2)^{1/2}$ (e.g. Binney & Tremaine 1987). The velocity dispersion viewed along the major

axis is larger than that on the minor, and the correction can be quite large since a typical galaxy with $q = 0.7$ will have a ratio $\sigma_{major}/\sigma_{minor} \simeq 1.16$ that is much larger than typical measurement uncertainties. If galaxies are oblate, this provides no help for the case of PG1115+080 because making the line-of-sight dispersion too high requires a prolate galaxy. However, it is a very simple means of shifting HST14176+5226. Crudely, if we start with the low 209 km/s velocity dispersion and assume that the lens is an $q = 0.7$ galaxy viewed pole on, then $\sigma_{major}/\sigma_{minor} \simeq 1.14$ and the corrections for the shape are large enough to make HST14176+5226 consistent with the other systems.

A final caveat is that neglecting necessary degrees of freedom in your lens model can bias inferences from the stellar dynamics of lenses just as it can in pure lens modeling. For example, Sand et al. (2002, 2003) used a comparison of lensed arcs in clusters to velocity dispersion measurements of the central cluster galaxy to argue that the cluster dark matter distribution could not have the $\rho \propto 1/r$ cusp of the NFW model for CDM halos. However, Bartelmann & Meneghetti (2003) and Dalal & Keeton (2003) show that the data are consistent with an NFW cusp if the lens models include a proper treatment of the non-spherical nature of the clusters. This has not been an issue in the stellar dynamics of strong lenses where the lens models used to determine the mass scale have always included the effects of ellipticity and shear, but it is well worth remembering.

B.5 Time Delays

Nothing compares to the measurement of the Hubble constant in bringing out the worst in astronomers. As we discussed in the previous section on lens modeling, many discussions of lens models seem obfuscatory rather than illuminating, and the tendency in this direction increases when the models are used to estimate H_0. In this section we discuss the relationship between time delay measurements, lens models and H_0. All results in the literature are consistent with this discussion, although it may take you several days and a series of e-mails to confirm it for any particular paper. The basic idea is simple. We see images at extrema of the virtual time delay surface (e.g. Blandford & Narayan 1986, Part 1) so the propagation time from the source to the observer differs for each image. The differences in propagation times, known as time delays, are proportional to H_0^{-1} because the distances between the observer, the lens and the source depend on the Hubble constant (Refsdal 1964). When the source varies, the variations appear in the images separated by the time delays and the delays are measured by cross-correlating the light curves. There are recent reviews of time delays and the Hubble constant by Courbin, Saha & Schechter (2002) and Kochanek & Schechter (2004). Portions of this section are adapted from Kochanek & Schechter (2004) since

we were completing that review at about the same time as we presented these lectures.

To begin the discussion we start with our standard simple model, the circular power law lens from §B.3. As a circular lens, we see two images at radii θ_A and θ_B from the lens center and we will assume that $\theta_A > \theta_B$ (Fig. B.20). Image A is a minimum, so source variability will appear in image A first and then with a time delay Δt in the saddle point image B. We can easily fit this data with an SIS lens model since (see Eqn. B.21 and B.22) to find that $\theta_A = \beta + b$ and $\theta_B = b - \beta$ where $b = (\theta_A + \theta_B)/2$ is the critical (Einstein) radius of the lens and $\beta = (\theta_A - \theta_B)/2$ is the source position. The light travel time for each image relative to a fiducial unperturbed ray is (see Part 1)

$$\tau(\boldsymbol{\theta}) = \frac{D_d D_s}{c D_{ds}} \left[\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \Psi(\boldsymbol{\theta}) \right] \tag{B.93}$$

where the effective potential $\Psi = b\theta$ for the SIS lens. Remember that the distances are comoving angular diameter distances rather than the more familiar angular diameter distances and this leads to the vanishing of the extra $1 + z_l$ factor that appears in the numerator if you insist on using angular diameter distances. The propagation time scales as $H_0^{-1} \simeq 10h^{-1}$ Gyr because of the H_0^{-1} scalings of the distances. After substituting our lens model, and differencing the delays for the two images, we find that

$$\Delta t_{SIS} = \tau_B - \tau_A = \frac{1}{2} \frac{D_d D_s}{c D_{ds}} \left(\theta_A^2 - \theta_B^2 \right). \tag{B.94}$$

The typical deflection angle $b \sim 3 \times 10^{-6}$ radians (so $R_A^2 \sim 10^{-11}$) converts the $10h^{-1}$ Gyr propagation time into a time delay of months or years that can be measured by a graduate student. Naively, this result suggests that the problem of interpreting time delays to measure H_0 is a trivial problem in astrometry.

We can check this assumption by using our general power-law models from §B.3 instead of an SIS. The power-law models correspond to density distributions $\rho \propto r^{-n}$, surface densities $\kappa \propto R^{1-n}$ and circular velocities $v_c \propto r^{(2-n)/2}$ of which the SIS model is the special case with $n = 2$. These models have effective potentials

$$\Psi(\boldsymbol{\theta}) = \frac{b^2}{3-n} \left(\frac{\theta}{b} \right)^{3-n}. \tag{B.95}$$

As we discussed in the §B.4.1 we can fit our simple, circular two-image lens with any of these models to determine $b(n)$ and $\beta(n)$ (Eqn. B.66), which we can then substitute into the expression for the propagation time to find that the time delay between the images is

$$\Delta t(n) = (n-1) \Delta t_{SIS} \left[1 - \frac{(2-n)^2}{12} \left(\frac{\delta\theta}{\langle\theta\rangle} \right)^2 + \cdots \right] \tag{B.96}$$

where we have expanded the result as a series in the ratio between the mean radius of the images $\langle\theta\rangle = (\theta_A+\theta_B)/2$ and the thickness of the radial annulus separating them $\delta\theta = \theta_A - \theta_B$. While the expansion assumes that $\delta\theta/\langle\theta\rangle \sim \beta/b$ is small, we can usually ignore higher order terms even when $\delta\theta/\langle\theta\rangle$ is of order unity. We now see that the time delay depends critically on the density profile, with more centrally concentrated mass distributions (larger values of n) producing longer time delays or implying larger Hubble constants for a fixed time delay.

The other idealization of the SIS model, the assumption of a circular lens, turns out to be less critical. A very nice analytic example is to consider a singular isothermal model with arbitrary angular structure in which $\kappa = bF(\chi)/2\theta$ where $F(\chi)$ is an arbitrary function of the azimuthal angle. The singular isothermal ellipsoid (Eqn. B.37) is an example of this class of potential. For this model family, $\Delta t = \Delta t_{SIS}$ independent of the actual angular structure $F(\chi)$ (Witt, Mao & Keeton 2000).

B.5.1 A General Theory of Time Delays

Just as for estimating mass distributions (§B.4), the aspect of modeling time delays that creates the greatest suspicion is the need to model the gravitational potential of the lens. Just as for mass distributions, this problem is largely of our own making, arising from poor communication, understanding and competition between groups. Here we will use simple mathematical expansions to show exactly what properties of the potential determine time delays. Any models which have these generic properties have all the degrees of freedom needed to properly interpret time delays. This does not, unfortunately, avoid the problem of degeneracies between the mass models and the Hubble constant.

The key to understanding time delays comes from Gorenstein, Falco & Shapiro (1988, Kochanek 2002a, see also Saha 2000) who showed that the time delay in a circular lens depends only on the image positions and *the surface density $\kappa(\theta)$ in the annulus between the images.* The two lensed images at radii $\theta_A > \theta_B$ define an annulus bounded by their radii, with an interior region for $\theta < \theta_B$ and an exterior region for $\theta > \theta_A$ (Fig. B.20). As we discussed in §B.4.1, the mass in the interior region is implicit in the image positions and constrained by the astrometry. From Gauss' law we know that the distribution of the mass in the interior and the amount or distribution of mass in the exterior region is irrelevant (see §B.4.3). A useful approximation is to assume that the surface density in the annulus can be *locally* approximated by a power law, $\kappa(\theta) \propto \theta^{1-n}$ for $\theta_B < \theta < \theta_A$, with a mean surface density in the annulus of $\langle\kappa\rangle = \langle\Sigma\rangle/\Sigma_c$. The time delay between the images is then (Kochanek 2002a)

$$\Delta t = 2\Delta t_{SIS}\left[1 - \langle\kappa\rangle - \frac{1-n\langle\kappa\rangle}{12}\left(\frac{\delta\theta}{\langle\theta\rangle}\right)^2 + O\left(\left(\frac{\delta\theta}{\langle\theta\rangle}\right)^4\right)\right] \qquad \text{(B.97)}$$

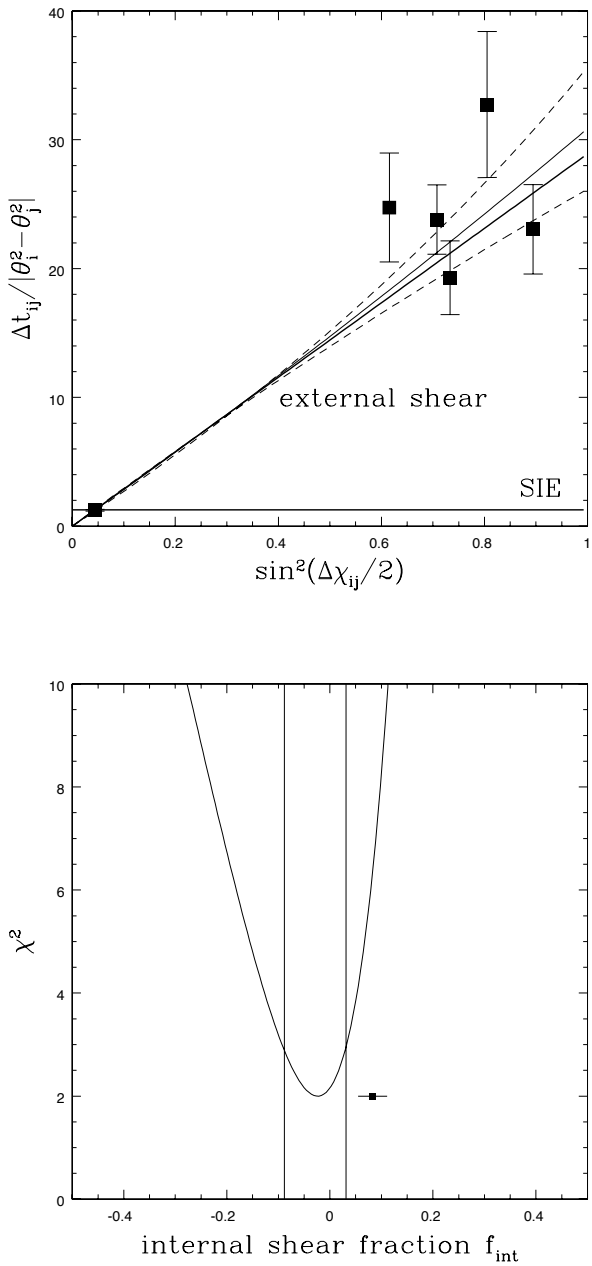


Fig. B.34. (Top) The PG1115+080 time delays scaled by the astrometric factor $\theta_i^2 - \theta_j^2$ appearing in Δt_{SIS} (Eqn. B.94) as a function of the leading angular dependence of the time delay ($\sin^2 \Delta\chi_{ij}/2$) (Eqn. B.98). The light solid curve and the dashed curves show the dependence for the best fit internal shear fraction f_{int} and its 68% confidence limits. A true external shear $f_{int} = 0$ is shown by the heavy solid curve inside the confidence limits, and the scaling for an SIE ($f_{int} = 1/4$) is shown by the horizontal line. (Bottom) The χ^2 goodness of fit for the internal shear fraction f_{int} from the time delay ratios is shown by the curve with the 68% confidence region bracketed by the vertical lines. The estimate of f_{int} from the image astrometry is shown by the point with an error bar.

where $\langle\theta\rangle = (\theta_A + \theta_B)/2$ and $\delta\theta = \theta_A - \theta_B$ as before. Thus, the time delay is largely determined by the average surface density $\langle\kappa\rangle$ in the annulus with only modest corrections from the local shape of the surface density distribution even when $\delta\theta/\langle\theta\rangle \sim 1$. This second order expansion is exact for an SIS lens ($\langle\kappa\rangle = 1/2$, $n = 2$), and it reproduces the time delay of a point mass lens ($\langle\kappa\rangle = 0$) to better than 1% even when $\delta\theta/\langle\theta\rangle = 1$. The local model also explains the scalings of the global power-law models. A $\kappa \propto \theta^{1-n}$ global power law has surface density $\langle\kappa\rangle = (3-n)/2$ near the Einstein ring, so the leading term of the time delay is $\Delta t = 2\Delta_{SIS}(1 - \langle\kappa\rangle) = (n-1)\Delta t_{SIS}$ just as in Eqn. B.96.

The role of the angular structure of the lens is easily incorporated into the expansion through the multipole expansion of §B.4. A quadrupole term in the potential with dimensionless amplitude ϵ_Ψ produces ray deflections of order $O(\epsilon_\Psi b)$ at the Einstein radius b of the lens. In a four-image lens, the quadrupole deflections are comparable to the fractional thickness of the annulus, $\epsilon_\Psi \simeq \delta\theta/\langle\theta\rangle$, while in a two-image lens they are smaller. For an ellipsoidal density distribution, the $\cos(2m\chi)$ multipole amplitude is smaller than the quadrupole amplitude by $\epsilon_{2m} \sim \epsilon_\Psi^m \lesssim (\delta\theta/\langle\theta\rangle)^m$. Hence, to lowest order in the expansion we only need to include the internal and external quadrupoles of the potential but not the changes of the quadrupoles in the annulus or any higher order multipoles. Remember that what counts is the angular structure of the potential rather than of the density, and that potentials are always much rounder than densities with a typical scaling of m^{-2}:m^{-1}:1 between the potential, deflections and surface density for the $\cos m\chi$ multipoles (see §B.4.4)

While the full expansion for independent internal and external quadrupoles is too complex to be informative, the leading term for the case when the internal and external quadrupoles are aligned is informative. We have an internal shear of amplitude Γ and an external shear of amplitude γ with $\chi_\gamma = \chi_\Gamma$ as defined in Eqns. B.51 and B.52. The leading term of the time delay is

$$\Delta t \simeq 2\Delta t_{SIS}\,(1 - \langle\kappa\rangle)\,\frac{\sin^2(\Delta\chi_{AB}/2)}{1 - 4f_{int}\cos^2(\Delta\chi_{AB}/2)} \tag{B.98}$$

where $\Delta\chi_{AB}$ is the angle between the images (Fig. B.20) and $f_{int} = \Gamma/(\Gamma + \gamma)$ is the internal quadrupole fraction we explored in Fig. B.31. We need not worry about a singular denominator – successful models of the image positions do not allow such configurations.

A two-image lens has too few astrometric constraints to fully constrain a model with independent, misaligned internal and external quadrupoles. Fortunately, when the lensed images lie on opposite sides of the lens galaxy ($\Delta\chi_{AB} \simeq \pi + \delta$ with $|\delta| \ll 1$), the time delay becomes insensitive to the quadrupole structure. Provided the angular deflections are smaller than the radial deflections ($|\delta|\langle\theta\rangle \lesssim \delta\theta$), the leading term of the time delay reduces to the result for a circular lens, $\Delta t = 2\Delta t_{SIS}(1 - \langle\kappa\rangle)$ if we minimize the

total shear of the lens. In the minimum shear solution the shear converges to the invariant shear (γ_1) and the other shear component $\gamma_2 = 0$ (see §B.4.5). If, however, you allow the other shear component to be non-zero, then you find that $\Delta t = 2\Delta t_{SIS}(1 - \langle\kappa\rangle - \gamma_2)$ to lowest order – the second shear component acts like a contribution to the convergence. In the absence of any other constraints, this adds a modest additional uncertainty (5–10%) to interpretations of time delays in two-image lenses. To first order its effects should average out in an ensemble of lenses because the extra shear has no preferred sign.

A four image lens has more astrometric constraints and can constrain a model with independent, misaligned internal and external quadrupoles – this was the basis of the Turner et al. (2004) summary of the internal to total quadrupole ratios shown in Fig. B.31. If the external shear dominates, then $f_{int} \simeq 0$ and the leading term of the delay becomes $\Delta t = 2\Delta t_{SIS}(1 - \langle\kappa\rangle)\sin^2 \Delta\chi_{AB}/2$. If the model is isothermal, like the $\Psi = \theta F(\chi)$ model we introduced in Eqn. B.42, then $f_{int} = 1/4$ and we obtain the Witt et al. (2000) result that the time delay is independent of the angle between the images $\Delta t \simeq 2\Delta t_{SIS}(1 - \langle\kappa\rangle)$. Thus, delay ratios in a four-image lens are largely determined by the angular structure and provide a check on the potential model. Unfortunately, the only lens with precisely measured delay ratios, B1608+656 (Fassnacht et al. 2002), also has two galaxies inside the Einstein ring and is a poor candidate for a simple multipole treatment (although it is dominated by an internal quadrupole as expected, see Fig. B.31). The delay ratios for PG1115+080 are less well measured (Schechter et al. 1997, Barkana 1997, Chartas 2003), but should be dominated by external shear since the estimate from the image astrometry is that $f_{int} = 0.083$ ($0.055 < f_{int} < 0.111$ at 95% confidence). Fig. B.34 shows the dependence of the PG1115+080 delays on the leading angular dependence of the time delay (Eqn. B.98) after scaling out the standard astrometry factor for the different radii of the images (Eqn. B.94). Formally, the estimate from the time delays that $f_{int} = -0.02$ ($-0.09 < f_{int} < 0.03$ at 68% confidence) is a little discrepant, but the two estimates agree at the 95% confidence level and there are still some systematic uncertainties in the shorter optical delays of PG1115+080. Changes in f_{int} between lenses is the reason Saha (2004) found significant scatter between time delays scaled only by Δt_{SIS}, since the time delay lenses range from external shear dominated systems like PG1115+080 to internal shear dominated systems like B1608+656.

B.5.2 Time Delay Lenses in Groups or Clusters

Most galaxies are not isolated, and many early-type lens galaxies are members of groups or clusters, so we need to consider the effects of the local environment on the time delays. Weak perturbations are easily understood since they will simply be additional contributions to the surface density (κ_c) and the external shear/quadrupole (γ_c) we discussed in §B.4. In general the

effects of the external shear γ_c are minimal because they either have little effect on the delays (two-image lenses) or are tightly constrained by either the astrometry or delay ratios (four-image lenses or systems with lensed host galaxies see §B.10). The problems arise from either the degeneracies associated with the surface density κ_c or the need for a complete, complicated cluster model.

The problem with κ_c is the infamous *mass-sheet degeneracy* (Part 1, Falco, Gorenstein & Shapiro 1985). If we have a model predicting a time delay Δt_0 and add a sheet of constant surface density κ_c, then the time delay is changed to $(1-\kappa_c)\Delta t_0$ without changing the image positions, flux ratios, or time delay ratios. Its effects can be understood from §B.5.1 as a contribution to the annular surface density with $\langle\kappa\rangle = \kappa_c$ and $\eta = 1$. Its only observable effect other than that on the time delays is a reduction in the mass of the lens galaxy that could be detected given an independent estimate of the lens galaxy's mass such as a velocity dispersion (e.g. §B.4.9 see Romanowsky & Kochanek 1998 for an attempt to to this for Q0957+561). It can also be done given an independent estimate of the properties of the group or cluster using weak lensing (e.g. Fischer et al. 1997 in Q0957+561), cluster galaxy velocity dispersions (e.g., Angonin-Willaime, Soucail, & Vanderriest 1994 for Q0957+561, Hjorth et al. 2002 for RXJ0911+0551) or X-ray temperatures/luminosities (e.g., Morgan et al. 2001 for RXJ0911+0551 or Chartas et al. 2002 for Q0957+561). The accuracy of these methods is uncertain at present because each suffers from its own systematic uncertainties, and they probably cannot supply the 10% or higher precision measurements of κ_c needed to strongly constrain models. When the convergence is due to an object like a cluster, there is a strong correlation between the convergence κ_c and the shear γ_c that is controlled by the density distribution of the cluster (for an isothermal model $\kappa_c = \gamma_c$). When the lens is in the outskirts of a cluster, as in RXJ0911+0551, it is probably reasonable to assume that $\kappa_c \leq \gamma_c$, as most mass distributions are more centrally concentrated than isothermal (see Eqn. B.8). Neglecting the extra surface density coming from nearby objects (galaxies, groups, clusters) leads to an overestimate of the Hubble constant, because these objects all have $\kappa_c > 0$. For most time delay systems this correction is probably $\lesssim 10\%$.

If the cluster or any member galaxies are sufficiently close, then we cannot ignore the higher-order perturbations in the expansion of Eqn. (B.26). This is certainly true for Q0957+561 (as discussed in §B.4.6) where the lens galaxy is the brightest cluster galaxy and located very close to the center of the cluster. It is easy to gauge when they become important by simply comparing the deflections produced by any higher order moments of the cluster beyond the quadrupole with the uncertainties being used for the image positions. For a cluster of critical radius b_c at distance θ_c from a lens of Einstein radius b, these perturbations are of order $b_c(b/\theta_c)^2 \sim b\gamma_c(b/\theta_c)$. Because the astrometric precision of the measurements is so high, these higher order terms can be

Table B.1. Time Delay Measurements

System	N_{im}	Δt (days)	Astrometry	Model	Ref.
HE1104–1805	2	161 ± 7	+	"simple"	1
PG1115+080	4	25 ± 2	+	"simple"	2
SBS1520+530	2	130 ± 3	+	"simple"	3
B1600+434	2	51 ± 2	+/−	"simple"	4
HE2149–2745	2	103 ± 12	+	"simple"	5
RXJ0911+0551	4	146 ± 4	+	cluster/satellite	6
Q0957+561	2	417 ± 3	+	cluster	7
B1608+656	4	77 ± 2	+/−	satellite	8
B0218+357	2	10.5 ± 0.2	−	"simple"	9
PKS1830–211	2	26 ± 4	−	"simple"	10
B1422+231	4	(8 ± 3)	+	"simple"	11

N_{im} is the number of images. Δt is the longest of the measured delays and its 1σ error; delays in parenthesis require further confirmation. The "Astrometry" column indicates the quality of the astrometric data for the system: + (good), +/− (some problems), − (serious problems). The "Model" column indicates the type of model needed to interpret the delays. "Simple" lenses can be modeled as a single primary lens galaxy in a perturbing tidal field. More complex models are needed if there is a satellite galaxy inside the Einstein ring ("satellite") of the primary lens galaxy, or if the primary lens belongs to a cluster. References: (1) Ofek & Maoz 2003, Wyrzykowski et al. 2003; (2) Barkana 1997, based on Schechter et al. 1997; (3) Burud et al. 2002b; (4) Burud et al. 2000, also Koopmans et al. 2000; (5) Burud et al. 2002a; (6) Hjorth et al. 2002; (7) Kundić et al. 1997, also Schild & Thomson 1997 and Haarsma et al. 1999; (8) Fassnacht et al. 2002; (9) Biggs et al. 1999, also Cohen et al. 2000; (10) Lovell et al. 1998; (11) Patnaik & Narasimha 2001.

relatively easy to detect. For example, models of PG1115+080 (e.g. Impey et al. 1998) find that using a group potential near the optical centroid of the nearby galaxies produces a better fit than simply using an external shear. In this case the higher order terms are fairly small and affect the results little, but results become very misleading if they are important but ignored.

B.5.3 Observing Time Delays and Time Delay Lenses

The first time delay measurement, for the gravitational lens Q0957+561, was reported in 1984 (Florentin-Nielsen 1984). Unfortunately, a controversy then developed between a short delay ($\simeq$ 1.1 years, Schild & Cholfin 1986; Vanderriest et al. 1989) and a long delay ($\simeq$ 1.5 years, Press, Rybicki, & Hewitt 1992a, 1992b), which was finally settled in favor of the short delay only after 5 more years of effort (Kundić et al. 1997; also Schild & Thomson 1997 and Haarsma et al. 1999). Factors contributing to the intervening difficulties included the small amplitude of the variations, systematic effects, which, with hindsight, appear to be due to microlensing and scheduling difficulties (both technical and sociological).

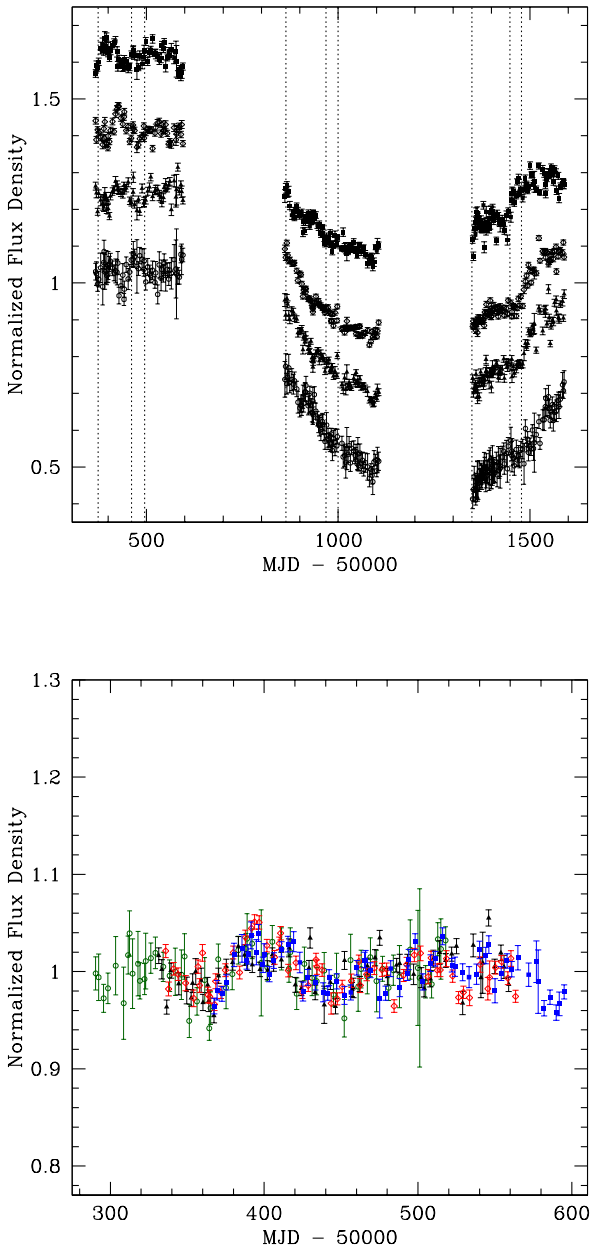


Fig. B.35. VLA monitoring data for the four-image lens B1608+656. The top panel shows (from top to bottom) the normalized light curves for the B (filled squares), A (open diamonds), C (filled triangles) and D (open circles) images as a function of the mean Julian day. The bottom panel shows the composite light curve for the first monitoring season after cross correlating the light curves to determine the time delays ($\Delta t_{AB} = 31.5 \pm 1.5$, $\Delta t_{CB} = 36.0 \pm 1.5$ and $\Delta t_{DB} = 77.0 \pm 1.5$ days) and the flux ratios. (From Fassnacht et al. 2002.)

While the long-running controversy over Q0957+561 led to poor publicity for the measurement of time delays, it allowed the community to come to an understanding of the systematic problems in measuring time delays, and to develop a broad range of methods for reliably determining time delays from typical data. Only the sociological problem of conducting large monitoring projects remains as an impediment to the measurement of time delays in large numbers. Even these are slowly being overcome, with the result that the last five years have seen the publication of time delays in 11 systems (see Table B.5.2).

The basic procedures for measuring a time delay are simple. A monitoring campaign must produce light curves for the individual lensed images that are well sampled compared to the time delays. During this period, the source quasar in the lens must have measurable brightness fluctuations on time scales shorter than the monitoring period. The resulting light curves are cross correlated by one or more methods to measure the delays and their uncertainties (e.g., Press et al. 1992a 1992b; Beskin & Oknyanskij 1995; Pelt et al. 1996; references in Table 1.1). Care must be taken because there can be sources of uncorrelated variability between the images due to systematic errors in the photometry and real effects such as microlensing of the individual images (e.g., Koopmans et al. 2000; Burud et al.2002b; Schechter et al. 2003). Figure B.35 shows an example, the beautiful light curves from the radio lens B1608+656 by Fassnacht et al. (2002), where the variations of all four lensed images have been traced for over three years. One of the 11 systems, B1422+231, is limited by systematic uncertainties in the delay measurements.

We want to have uncertainties in the time delay measurements that are unimportant for the estimates of H_0. For the present, uncertainties of order 3%–5% are adequate (so improved delays are still needed for PG1115+080, HE2149–2745, and PKS1830–211). In a four-image lens we can measure three independent time delays, and the dimensionless ratios of these delays provide additional constraints on the lens models (see §B.5.1). These ratios are well measured in B1608+656 (Fassnacht et al. 2002), poorly measured in PG1115+080 (Barkana 1997; Schechter et al. 1997; Chartas 2003) and unmeasured in either RXJ0911+0551 or B1422+231. Using the time delay lenses as very precise probes of H_0, dark matter and cosmology will eventually require still smaller delay uncertainties ($\sim 1\%$). Once a delay is known to 5%, it is relatively easy to reduce the uncertainties further because we can accurately predict when flux variations will appear in the other images and the lens will need to be monitored more intensively.

The expression for the time delay in an SIS lens (Eqn. B.94) reveals the other data that are necessary to interpret time delays. First, the source and lens redshifts are needed to compute the distance factors that set the scale of the time delays. Fortunately, we know both redshifts for all 11 systems in Table B.5.2 even though missing redshifts are a problem for the lens sample as a whole (see §B.2). The dependence of the distances D_d, D_s and D_{ds} on

the cosmological model is unimportant until our total uncertainties approach 5%. Second, we require accurate relative positions for the images and the lens galaxy. These uncertainties are always dominated by the position of the lens galaxy relative to the images. For most of the lenses in Table B.5.2, observations with radio interferometers (VLA, Merlin, VLBA) and HST have measured the relative positions of the images and lenses to accuracies $\lesssim 0\farcs005$. Sufficiently deep HST images can obtain the necessary data for almost any lens, but dust in the lens galaxy (as seen in B1600+434 and B1608+656) can limit the accuracy of the measurement even in a very deep image. For B0218+357 and PKS1830–211, however, the position of the lens galaxy relative to the images is not known to sufficient precision or determined only from models (see Biggs et al. 1999, Léhar et al. 2000; Courbin et al. 2002; Winn et al. 2002b, Wucknitz, Biggs & Browne 2004, York et al. 2004).

We can also divide the systems by the complexity of the required lens model. We define eight of the lenses as "simple," in the sense that the available data suggests that a model consisting of a single primary lens in a perturbing shear (tidal gravity) field should be an adequate representation of the gravitational potential. In some of these cases, an external potential representing a nearby galaxy or parent group will improve the fits, but the differences between the tidal model and the more complicated perturbing potential are small (see §B.5.2). If we neglect the convergence produced by the group, then H_0 may be overestimated. We include the quotation marks because the classification is based on an impression of the systems from the available data and models. Remember also that there are convergence fluctuations along the line of sight that add a low level of cosmic variance to the time delays of individual lenses (§B.4.2 and Fig. B.19). While we cannot guarantee that a system is simple, we can easily recognize two complications that will require more complex models.

The first complication is that some primary lenses have less massive satellite galaxies inside or near their Einstein rings. This includes two of the time delay lenses, RXJ0911+0551 and B1608+656. RXJ0911+0551 could simply be a projection effect, since neither lens galaxy shows irregular isophotes. Here the implication for models may simply be the need to include all the parameters (mass, position, ellipticity $\cdots$) required to describe the second lens galaxy, and with more parameters we would expect greater uncertainties in H_0. In B1608+656, however, the lens galaxies show the disturbed isophotes of dusty galaxies possibly undergoing a disruptive interaction. How one should model such a system is unclear. If there was once dark matter associated with each of the galaxies, how is it distributed now? Is it still associated with the individual galaxies? Has it settled into an equilibrium configuration? While B1608+656 can be well fit with standard lens models (Fassnacht et al. 2002, Koopmans et al. 2003), these complications have yet to be explored in detail.

The second complication occurs when the primary lens is a member of a more massive (X-ray) cluster, as in the time delay lenses RXJ0911+0551

(Morgan et al. 2001) and Q0957+561 (Chartas et al. 2002). The cluster model is critical to interpreting these systems (see §B.5.2). The cluster surface density at the position of the lens ($\kappa_c \gtrsim 0.2$) leads to large corrections to the time delay estimates and the higher-order perturbations are crucial to obtaining a good model. For example, models in which the Q0957+561 cluster was treated simply as an external shear were grossly incorrect (see §B.4.6, Keeton et al. 2000). In addition to the uncertainties in the cluster model itself, we must also decide how to include and model the other cluster galaxies near the primary lens. Thus, lenses in clusters have many reasonable degrees of freedom beyond those of the "simple" lenses.

B.5.4 Results: The Hubble Constant and Dark Matter

With our understanding of the theory and observations of the lenses we will now explore their implications for H_0. We focus on the "simple" lenses PG1115+080, SBS1520+530, B1600+434, and HE2149–2745. We only comment on the interpretation of the HE1104–1805 delay because the measurement is too recent to have been interpreted carefully. We will briefly discuss the more complicated systems B0218+357, RXJ0911+0551, Q0957+561, and B1608+656, and we will not discuss the systems with problematic time delays or astrometry.

The most common, simple, realistic model of a lens consists of a singular isothermal ellipsoid (SIE) in an external (tidal) shear field (see §B.4). The model has 7 parameters (the lens position, mass, ellipticity, major axis orientation for the SIE, and the shear amplitude and orientation). It has many degrees of freedom associated with the angular structure of the potential, but the radial structure is fixed with $\langle\kappa\rangle \simeq 1/2$. For comparison, a two-image (four-image) lens supplies 5 (13) constraints on any model of the potential: 2 (6) from the relative positions of the images, 1 (3) from the flux ratios of the images, 0 (2) from the inter-image time delay ratios, and 2 from the lens position. With the addition of extra components (satellites/clusters) for the more complex lenses, this basic model provides a good fit to all the time delay lenses except Q0957+561. Although a naive counting of the degrees of freedom ($N_{dof} = -2$ and 6, respectively) suggests that estimates of H_0 would be under constrained for two-image lenses and over constrained for four-image lenses, the uncertainties are actually dominated by those of the time delay measurements and the astrometry in both cases. This is what we expect from §B.5.1 — the model has no degrees of freedom that change $\langle\kappa\rangle$ or η, so there will be little contribution to the uncertainties in H_0 from the model for the potential.

If we use a model that includes parameters to control the radial density profile (i.e., $\langle\kappa\rangle$), for example by adding a halo truncation radius a to the SIS profile [the pseudo-Jaffe model, $\rho \propto r^{-2}(r^2 + a^2)^{-1}$; e.g., Impey et al. 1998;

Burud et al. 2002a],[5] then we find the expected correlation between a and H_0 — as we make the halo more concentrated (smaller a), the estimate of H_0 rises from the value for the SIS profile ($\langle\kappa\rangle = 1/2$ as $a \to \infty$) to the value for a point mass ($\langle\kappa\rangle = 0$ as $a \to 0$), with the fastest changes occurring when a is similar to the Einstein radius of the lens. We show an example of such a model for PG1115+080 in Figure B.36. This case is somewhat more complicated than a pure pseudo-Jaffe model because there is an additional contribution to the surface density from the group to which the lens galaxy belongs. As long as the structure of the radial density profile is fixed (constant a), the uncertainties are again dominated by the uncertainties in the time delay. Unfortunately, the goodness of fit, $\chi^2(a)$, shows too little dependence on a to determine H_0 uniquely. In general, two-image lenses have too few constraints, and the extra constraints supplied by a four-image lens constrain the angular structure rather than the radial structure of the potential. This basic problem holds for all existing models of the current sample of time delay lenses.

The inability of the present time delay lenses to directly constrain the radial density profile is the major problem for using them to determine H_0. Fortunately, it is a consequence of the available data on the current sample rather than a fundamental limitation. It is, however, a simple trade-off – models with less dark matter (lower $\langle\kappa\rangle$, more centrally concentrated densities) produce higher Hubble constants than those with more dark matter. We do have some theoretical limits on the value of $\langle\kappa\rangle$. In particular, we can be confident that the surface density is bounded by two limiting models. The mass distribution should not be more compact than the luminosity distribution, so a constant mass-to-light ratio (M/L) model should set a lower limit on $\langle\kappa\rangle \gtrsim \langle\kappa\rangle_{M/L} \simeq 0.2$, and an upper limit on estimates of H_0. We are also confident that the typical lens should not have a rising rotation curve at 1–2 optical effective radii from the center of the lens galaxy. Thus, the SIS model is probably the least concentrated reasonable model, setting an upper bound on $\langle\kappa\rangle \lesssim \langle\kappa\rangle_{SIS} = 1/2$, and a lower limit on estimates of H_0. Figure B.37 shows joint estimates of H_0 from the four simple lenses for these two limiting mass distributions (Kochanek 2003b). The results for the individual lenses are mutually consistent and are unchanged by the new 0.149 ± 0.004 day delay for the A_1-A_2 images in PG1115+080 (Chartas 2003). For galaxies with isothermal profiles we find $H_0 = 48 \pm 3$ km s^{-1} Mpc^{-1}, and for galaxies with constant M/L we find $H_0 = 71 \pm 3$ km s^{-1} Mpc^{-1}. While our best prior estimate for the mass distribution is the isothermal profile (see §B.4.6), the lens galaxies would have to have constant M/L to match Key Project estimate of $H_0 = 72 \pm 8$ km s^{-1} Mpc^{-1} (Freedman et al. 2001) or the WMAP estimate of $H_0 = 72 \pm 5$ km s^{-1} Mpc^{-1} for a flat universe with a cosmological constant (Spergel et al. 2003).

[5] This is simply an example. The same behavior would be seen for any other parametric model in which the radial density profile can be adjusted.

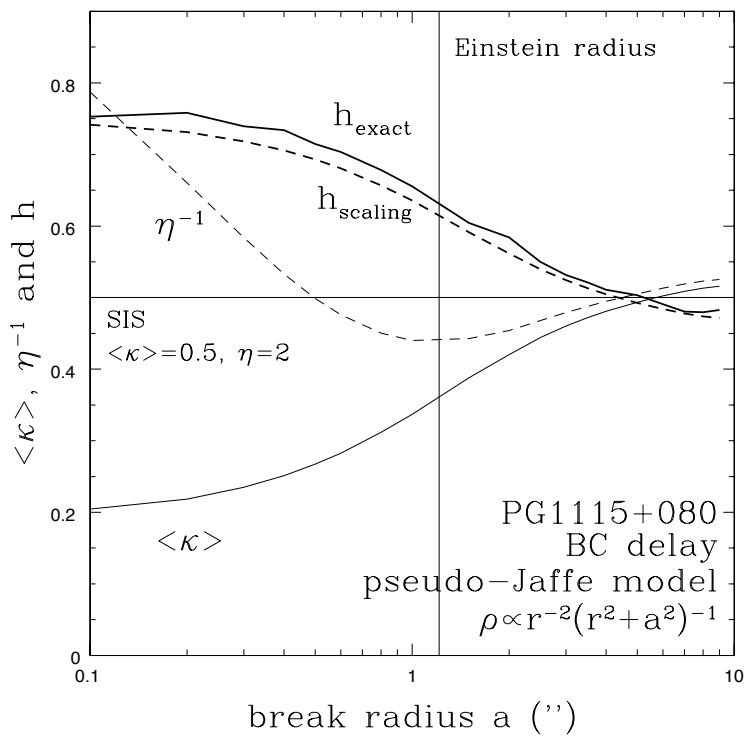


Fig. B.36. H_0 estimates for PG1115+080. The lens galaxy is modeled as an ellipsoidal pseudo-Jaffe model, $\rho \propto r^{-2}(r^2+a^2)^{-1}$, and the nearby group is modeled as an SIS. As the break radius $a \to \infty$ the pseudo-Jaffe model becomes an SIS model, and as the break radius $a \to 0$ it becomes a point mass. The heavy solid curve (h_{exact}) shows the dependence of H_0 on the break radius for the exact, nonlinear fits of the model to the PG1115+080 data. The heavy dashed curve ($h_{scaling}$) is the value found using our simple theory (§B.5.1) of time delays. The agreement of the exact and scaling solutions is typical. The light solid line shows the average surface density $\langle\kappa\rangle$ in the annulus between the images, and the light dashed line shows the *inverse* of the logarithmic slope η in the annulus ($\kappa \propto \theta^{1-\eta}$). For an SIS model we would have $\langle\kappa\rangle = 1/2$ and $\eta^{-1} = 1/2$, as shown by the horizontal line. When the break radius is large compared to the Einstein radius (indicated by the vertical line), the surface density is slightly higher and the slope is slightly shallower than for the SIS model because of the added surface density from the group. As we make the lens galaxy more compact by reducing the break radius, the surface density decreases and the slope becomes steeper, leading to a rise in H_0. As the galaxy becomes very compact, the surface density near the Einstein ring is dominated by the group rather than the galaxy, so the surface density approaches a constant and the logarithmic slope approaches the value corresponding to a constant density sheet ($\eta = 1$).

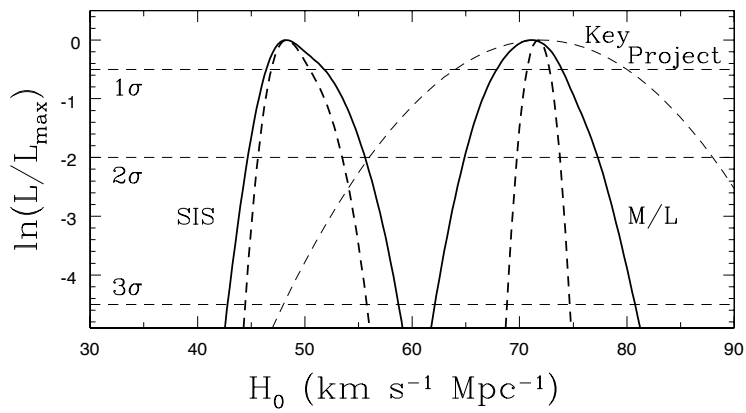


Fig. B.37. H_0 likelihood distributions. The curves show the joint likelihood functions for H_0 using the four simple lenses PG1115+080, SBS1520+530, B1600+434, and HE2149–2745 and assuming either an SIS model (high $\langle\kappa\rangle$, flat rotation curve) or a constant M/L model (low $\langle\kappa\rangle$, declining rotation curve). The heavy dashed curves show the consequence of including the X-ray time delay for PG1115+080 from Chartas (2003) in the models. The light dashed curve shows a Gaussian model for the Key Project result that $H_0 = 72 \pm 8$ km s^{-1} Mpc^{-1}.

The difference between these two limits is entirely explained by the differences in $\langle\kappa\rangle$ and η between the SIS and constant M/L models. In fact, it is possible to reduce the H_0 estimates for each simple lens to an approximation formula, $H_0 = A(1-\langle\kappa\rangle) + B\langle\kappa\rangle(\eta-1)$. The coefficients, A and $|B| \approx A/10$, are derived from the image positions and the time delay using the simple theory from §B.5.1. These approximations reproduce numerical results using ellipsoidal lens models to accuracies of 3 km s^{-1} Mpc^{-1} (Kochanek 2002a). For example, in Figure B.36 we also show the estimate of H_0 computed based on the simple theory of §B.5.1 and the annular surface density ($\langle\kappa\rangle$) and slope (η) of the numerical models. The agreement with the full numerical solutions is excellent, even though the numerical models include both the ellipsoidal lens galaxy and a group. No matter what the mass distribution is, the five lenses PG1115+080, SBS1520+530, B1600+434, PKS1830–211,[6] and HE2149–2745 have very similar dark matter halos. For a fixed slope η, the five systems are consistent with a common value for the surface density of

$$\langle\kappa\rangle = 1 - 1.07h + 0.14(\eta-1)(1-h) \pm 0.04 \tag{B.99}$$

[6] PKS1830–211 is included based on the Winn et al. (2002b) model of the *HST* imaging data as a single lens galaxy. Courbin et al. (2002) prefer an interpretation with multiple lens galaxies which would invalidate the analysis.

where $H_0 = 100h$ km s^{-1} Mpc^{-1} and there is an upper limit of $\sigma_\kappa \lesssim 0.07$ on the intrinsic scatter of $\langle\kappa\rangle$. Thus, time delay lenses provide a new window into the structure and homogeneity of dark matter halos, regardless of the actual value of H_0.

There is an enormous range of parametric models that can illustrate how the extent of the halo affects $\langle\kappa\rangle$ and hence H_0 — the pseudo-Jaffe model we used above is only one example. It is useful, however, to use a physically motivated model where the lens galaxy is embedded in a standard NFW (Navarro, Frenk, & White 1996) profile halo as we discussed at the end of §B.4.1. The lens galaxy consists of the baryons that have cooled to form stars, so the mass of the visible galaxy can be parameterized using the cold baryon fraction $f_{b,cold}$ of the halo, and for these CDM halo models the value of $\langle\kappa\rangle$ is controlled by the cold baryon fraction (Kochanek 2003a). A constant M/L model is the limit $f_{b,cold} \to 1$ (with $\langle\kappa\rangle \simeq 0.2$, $\eta \simeq 3$). Since the baryonic mass fraction of a CDM halo should not exceed the global fraction of $f_b \simeq 0.17 \pm 0.03$ (e.g., Spergel et al. 2003), we cannot use constant M/L models without also abandoning CDM. As we reduce $f_{b,cold}$, we are adding mass to an extended halo around the lens, leading to an increase in $\langle\kappa\rangle$ and a decrease in η. For $f_{b,cold} \simeq 0.02$ the model closely resembles the SIS model ($\langle\kappa\rangle \simeq 1/2$, $\eta \simeq 2$). If we reduce $f_{b,cold}$ further, the mass distribution begins to approach that of the NFW halo without any cold baryons. Figure B.38 shows how $\langle\kappa\rangle$ and H_0 depend on $f_{b,cold}$ for PG1115+080, SBS1520+530, B1600+434 and HE2149–2745. When $f_{b,cold} \simeq 0.02$, the CDM models have parameters very similar to the SIS model, and we obtain a very similar estimate of $H_0 = 52 \pm 6$ km s^{-1} Mpc^{-1} (95% confidence). If all baryons cool, and $f_{b,cold} = f_b$, then we obtain $H_0 = 65 \pm 6$ km s^{-1} Mpc^{-1} (95% confidence), which is still lower than the Key Project estimates.

We excluded the lenses requiring significantly more complicated models with multiple lens galaxies or very strong perturbations from clusters. If we have yet to reach a consensus on the mass distribution of relatively isolated lenses, it seems premature to extend the discussion to still more complicated systems. We can, however, show that the clusters lenses require significant contributions to $\langle\kappa\rangle$ from the cluster in order to produce the same H_0 as the more isolated systems. As we discussed in §B.2 the three more complex systems are RXJ0911+0551, Q0957+561 and B1608+656.

RXJ0911+0551 is very strongly perturbed by the nearby X-ray cluster (Morgan et al. 2001; Hjorth et al. 2002). Kochanek (2003b) found $H_0 = 49 \pm 5$ km s^{-1} Mpc^{-1} if the primary lens and its satellite were isothermal and $H_0 = 67 \pm 5$ km s^{-1} Mpc^{-1} if they had constant mass-to-light ratios. The higher value of $H_0 = 71 \pm 4$ km s^{-1} Mpc^{-1} obtained by Hjorth et al. (2002) can be understood by combining §B.5.1 and §B.5.2 with the differences in the models. In particular, Hjorth et al. (2002) truncated the halo of the primary lens near the Einstein radius and used a lower mass cluster, both of which

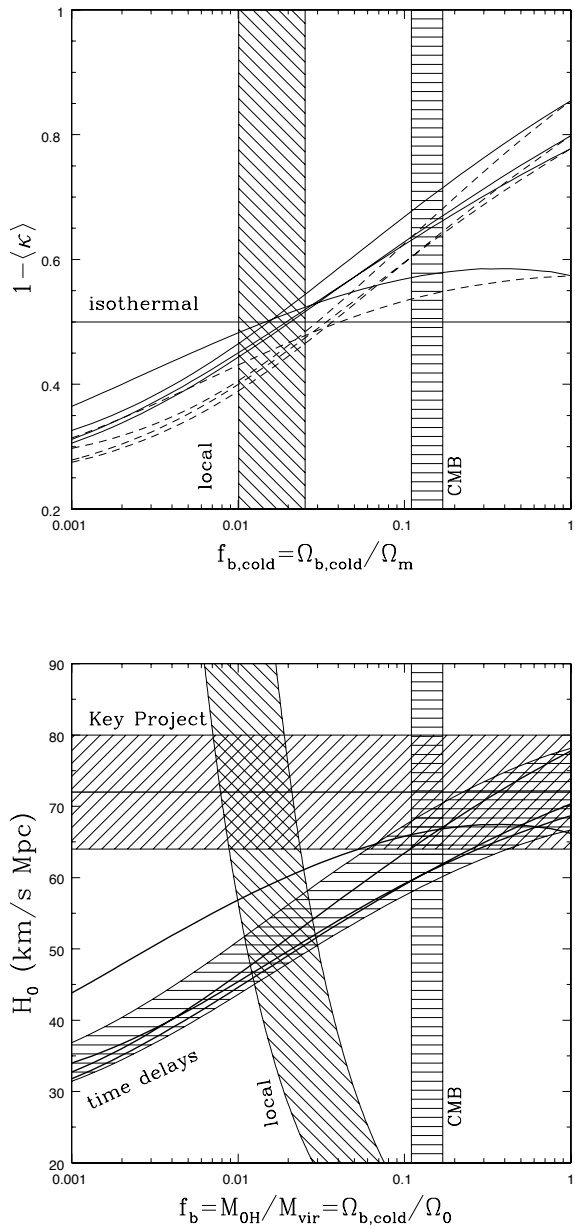


Fig. B.38. H_0 in CDM halo models. The top panel shows $1-\langle\kappa\rangle$ for the "simple" lenses (PG1115+080, SBS1520+530, B1600+434, and HE2149–2745) as a function of the cold baryon fraction $f_{b,cold}$. The solid (dashed) curves include (exclude) the adiabatic compression of the dark matter by the baryons. The horizontal line shows the value for an SIS potential. The bottom panel shows the resulting estimates of H_0, where the shaded envelope bracketing the curves is the 95% confidence region for the combined lens sample. The horizontal band shows the Key Project estimate. For larger $f_{b,cold}$, the density $\langle\kappa\rangle$ decreases and the local slope η steepens, leading to larger values of H_0. The vertical bands in the two panels show the lower bound on f_b from local inventories and the upper bound from the CMB.

lower $\langle\kappa\rangle$ and raise H_0. The Hjorth et al. (2002) models also included many more cluster galaxies assuming fixed masses and halo sizes.

Q0957+561 is a special case because the primary lens galaxy is the brightest cluster galaxy and it lies nearly at the cluster center (Keeton et al. 2000; Chartas et al. 2002). As a result, the lens modeling problems are particularly severe, and Keeton et al. (2000) found that all previous models (most recently, Barkana et al. 1999; Bernstein & Fischer 1999; and Chae 1999, see §B.4.6) were incompatible with the observed geometry of the lensed host galaxy. While Keeton et al. (2000) found models consistent with the structure of the lensed host, they covered a range of almost $\pm 25\%$ in their estimates of H_0. A satisfactory treatment of this lens remains elusive.

HE1104–1805 has the most recently measured time delay(Ofek & Maoz 2003, Wyrzykowski et al. 2003). Given the $\Delta t = 161 \pm 7$ day delay, a standard SIE model of this system predicts a very high $H_0 \simeq 90$ km s^{-1} Mpc^{-1}. The geometry of this system and the fact that the inner image is brighter than the outer image both suggest that HE1104–1805 lies in an anomalously high tidal shear field, while the standard model includes a prior to keep the external shear small. A prior is needed because a two-image lens supplies too few constraints to determine both the ellipticity of the main lens and the external shear simultaneously. Since the images and the lens in HE1104–1805 are nearly collinear, the anomalous H_0 estimate for the standard model may be an example of the shear degeneracy we briefly mentioned in §B.5.1. At present the model surveys needed to understand the new delay have not been made. Observations of the geometry of the host galaxy Einstein ring will resolve any ambiguities due to the shear in the near future (see §B.10).

The lens B1608+656 consists of two interacting galaxies, and, as we discussed in §B.2, this leads to a greatly increased parameter space. Fassnacht et al. (2002) used SIE models for the two galaxies to find $H_0 = 61 - 65$ km s^{-1} Mpc^{-1}, depending on whether the lens galaxy positions are taken from the H-band or I-band lens *HST* images (the statistical errors are negligible). The position differences are probably created by extinction effects from the dust in the lens galaxies. Like isothermal models of the "simple" lenses, the H_0 estimate is below local values, but the disagreement is smaller. These models correctly match the observed time delay ratios. Koopmans et al. (2003) obtain a still higher estimate of $H_0 = 75 \pm 7$ km s^{-1} Mpc^{-1} largely because the lens galaxy positions shift after they include extinction corrections. They use a foreground screen model to make the extinction corrections, which is a better approximation than no extinction corrections, but is unlikely allow precise correction in a system like B1608+656 where the dust and stars are mixed and there is no simple relation between color excess and optical depth (e.g. Witt, Thronson & Capuano 1992).

Despite recent progress both in modeling the VLBI structure (Wucknitz et al. 2004) and obtaining deep images (York et al. 2004) it is unclear whether B0218+357 has escaped its problems with astrometry and models. While

York et al. (2004) have clearly measured the position of the lens galaxy, the dependence of the position on the choice of the PSF model remains a significant source of uncertainty for estimates of H_0. Models of the system using power law models find a slope very close to isothermal $\eta = 2.04 \pm 0.02$ ($\rho \propto r^{-\eta}$). Unfortunately, these models have too few degrees of freedom given the small astrometric uncertainties in the VLBI structures providing the constraints (because the only angular structure in the model is the ellipsoidal potential used for the main lens galaxy), and this makes the limits on the power slope suspect (see §B.4.6). For example, while it is true that Lehár et al. (2000) estimated that the environmental shear near B0218+357 was small, even a $\gamma = 0.01$ external tidal shear produces deflections (3 milliarcseconds) that are large compared to the accuracy of the constraints used for the models and so must be included for the models to be reliable. With these caveats, B0218+357 (like the models of B1608+656 with significant extinction corrections) support a nearly isothermal mass distribution with $H_0 = 73 \pm 8$ km s^{-1} Mpc^{-1}.

B.5.5 The Future of Time Delay Measurements

We understand the theory of time delays very well – the only important variable in the lens structure is the average surface density $\langle\kappa\rangle$ of the lens near the images for which the delay is measured. The angular structure of the potential has an effect on the delays, but it is either small or well-constrained by the observed image positions. Provided a lens does not lie in a cluster where the cluster potential cannot be described by a simple expansion, any lens model that includes the parameters needed to vary the average surface density of the lens near the images and to change the ratio between the quadrupole moment of the lens and the environment includes all the variables needed to model time delays, to estimate the Hubble constant, and to understand the systematic uncertainties in the results. Unfortunately, there is a tendency in the literature to confuse rather than to illuminate this understanding, even though all differences between estimates of the Hubble constant for the simple time delay lenses can be understood on this basis.

The problem with time delays lies with the confusing state of the data. The four simplest time delay lenses, PG1115+080, SBS1520+530, B1600+434 and HE2149–2745, can only match the currently preferred estimate of $H_0 \simeq 72 \pm 8$ km s^{-1} Mpc^{-1} (Freedman et al. 2001, Spergel et al. 2003) if they have nearly constant M/L mass distributions. If they have the favored quasi-isothermal mass distributions, then $H_0 \simeq 48 \pm 3$ km s^{-1} Mpc^{-1}. This leads to a conundrum: why do simple lenses with time delay measurements have falling rotation curves, while simple lenses with direct estimates of the mass profile do not? This is further confused by B1608+656 and B0218+357, which due to their observational complexity would be the last systems I would attempt to understand, but in current analyses can be both isothermal and have high H_0. In resolving this problem it is not enough to search for a "Golden Lens."

There is no such thing! While chanting "My lens is better than your lens!" may be satisfying, it contributes little to understanding the basic problem.

The difficulty at the moment is that systems I would view as problematic (B0218+357 due to problems in astrometry or B1608+656 due to the interacting lens galaxies) allow both mass distributions with flat rotation curves and $H_0 = 72$ km s^{-1} Mpc^{-1}, while systems that should be simpler to interpret (the simple lenses in Table B.5.2) do not. Yet the preponderance of evidence on the mass distributions of lens galaxies suggests that they are fairly homogeneous in structure and have roughly flat rotation curves (§B.4). The simplest way to clarify this problem is to measure accurate time delays for many more systems. At a fixed value of the Hubble constant we will either find significant scatter in the surface densities near the images of simple lenses or we will not.

B.6 Gravitational Lens Statistics

It is the opinion of the author that the statistics of lenses as a method for determining the cosmological model has largely ceased to be interesting. However, it is important to understand the underlying physics because it determines the types of lenses we detect. While most recent analyses have found cosmological results consistent with the concordance model (Chae et al. 2002, Chae 2003, Davis, Huterer & Krauss 2003, Mitchell et al. 2004) there are still large statistical uncertainties and some dangerous systematic assumptions. More importantly, there is little prospect at present of lens statistics becoming competitive with other methods. Gravitational lenses statistics arguably begins with Press & Gunn (1973), although the "modern" era begins with the introduction of magnification bias (Turner 1980), the basic statistics of normal galaxy lenses (Turner, Ostriker & Gott 1984), cross sections and optical depths for more general lenses (Blandford & Kochanek 1987a, Kochanek & Blandford 1987), explorations of the effects of general cosmologies (Fukugita et al. 1990, Fukugita & Turner 1991) and lens structure (Maoz & Rix 1993, Kochanek 1996a) and the development of the general methodology of interpreting observations (Kochanek 1993b, 1996a).

B.6.1 The Mechanics of Surveys

There are two basic approaches to searching for gravitational lenses depending on whether you start with a list of potentially lensed sources or a list of potential lens galaxies. Of the two, only a search of sources for lensed sources has a significant cosmological sensitivity – for a non-evolving population of lenses in a flat cosmological model we will find in §B.6.3 that the number of lensed sources scales with the volume between the observer and the source D_s^3. If you search potential lens galaxies for those which have actually lensed a source, then the cosmological dependence enters only through distance ratios,

D_{ds}/D_s, and you require a precise knowledge of the source redshift distribution. Thus, while lenses found in this manner are very useful for many projects (mass distributions, galaxy evolution etc.), they are not very useful for determining the cosmological model. This changes for the case of cluster lenses where you may find multiple lensed sources at different redshifts behind the same lens (e.g. Soucail, Kneib & Gorse 2004).

Most lenses have been found by searching for lensed sources because the number of targets which must be surveyed is considerably smaller. This is basically a statement about the relative surface densities of candidate sources and lenses. The typical lens is a galaxy with an Einstein radius of approximately $b \simeq 1\farcs0$ so it has a cross section of order πb^2. If you search N lenses with such a cross section for signs of a lensed source, you would expect to find $N\pi b^2 \Sigma_{source}$ lenses where Σ_{source} is the surface density of detectable sources. If you search N sources for a lens galaxy in front of them, you would expect to find $N\pi b^2 \Sigma_{lens}$ lenses, where Σ_{lens} is the surface density of lens galaxies. Since the surface density of massive galaxies is significantly higher than the surface density of easily detectable higher redshift sources ($\Sigma_{lens} \gg \Sigma_{source}$), you need examine fewer sources than lens galaxies to find the same number of lensed systems. This is somewhat mitigated by the fact the surface density of potential lens galaxies is high enough to allow you to examine many potential lenses in a single observation, while the surface density of sources is usually so low that they can be examined only one at a time.

For these reasons, we present a short synopsis of searches for sources behind lenses and devote most of this section to the search for lenses in front of sources. The first method for finding sources behind lenses is a simple byproduct of redshift surveys. Redshift surveys take spectra of the central regions of low redshift galaxies allowing the detection of spectral features from any lensed images inside the aperture used for the spectrum. Thus, the lens Q2237+0305 was found in the CfA redshift survey (Huchra et al. 1985) and SDSS0903+5028 (Johnston et al. 2003) was found in the SDSS survey. Theoretical estimates (Kochanek 1992b, Mortlock & Webster 2000c) suggest that the discovery rate should one per 10^4–10^5 redshift measurements, but this does not seem to be borne out by the number of systems discovered in this age of massive redshift surveys (the origin of the lower rate in the 2dF survey is discussed by Mortlock & Webster 2001). Miralda-Escude & Lehár (1992) proposed searching for lensed optical (emission line) rings, a strategy successfully used by Warren et al. (1996) to find 0047–2808 and by Ratnatunga, Griffiths & Ostrander (1999) to find lenses in the HST Medium Deep Survey (MDS). There is also a hybrid approach whose main objective is simply to find lenses with minimal follow up observations by looking for high redshift radio lobes that have non-stellar optical counterparts (Lehár et al. 2001). Since radio lobes have no intrinsic optical emission, a lobe superposed on a galaxy is an excellent lens candidate. The present limitation on this method is the low angular resolution of the available all sky radio sur-

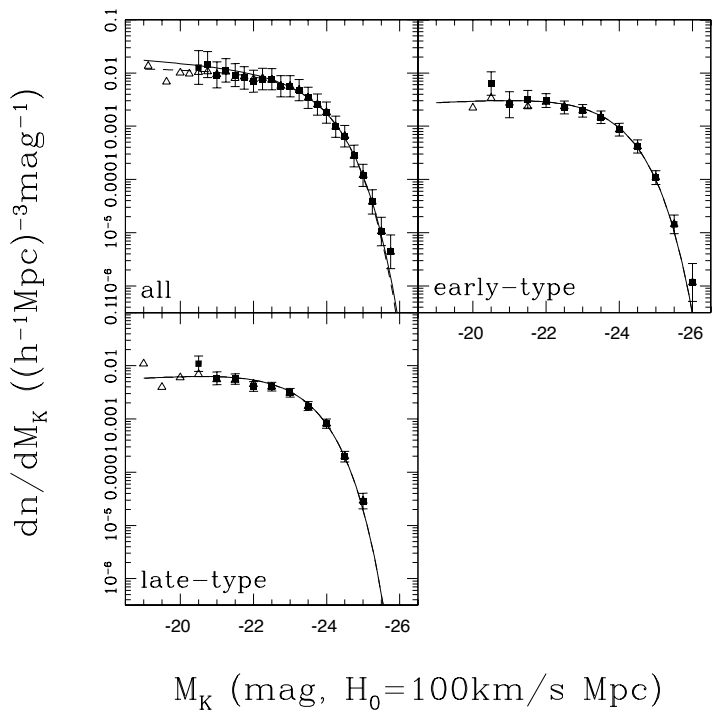


Fig. B.39. Example of a local galaxy luminosity function. These are the K-band luminosity functions for either all galaxies or by morphological type from Kochanek et al. (2001). The curves show the best fit Schechter models for the luminosity functions while the points with error bars show a non-parametric reconstruction.

veys (FIRST, NVSS) and the magnitude limits and star/galaxy separation problems of the current all-sky optical catalogs. Nonetheless, several systems have been discovered by this technique.

The majority of lens surveys, however, have focused on either optical quasars or radio sources because they are source populations known to lie at relatively high redshift ($z_s \gtrsim 1$) and that are easily detected even when there is an intervening lens galaxy. Surveys of optical quasars (Crampton, McClure & Fletcher 1992, Yee, Fillipenko & Tang 1993, Maoz et al. 1993, Surdej et al. 1993, Kochanek, Falco & Schild 1995) have the advantage that the sources are bright, and the disadvantages that the bright sources can mask the lens galaxy and that the selection process is modified by dust in the lens galaxy and emission from the lens galaxy. We will discuss these effects in §B.9. While many more lensed quasars have been discovered since these

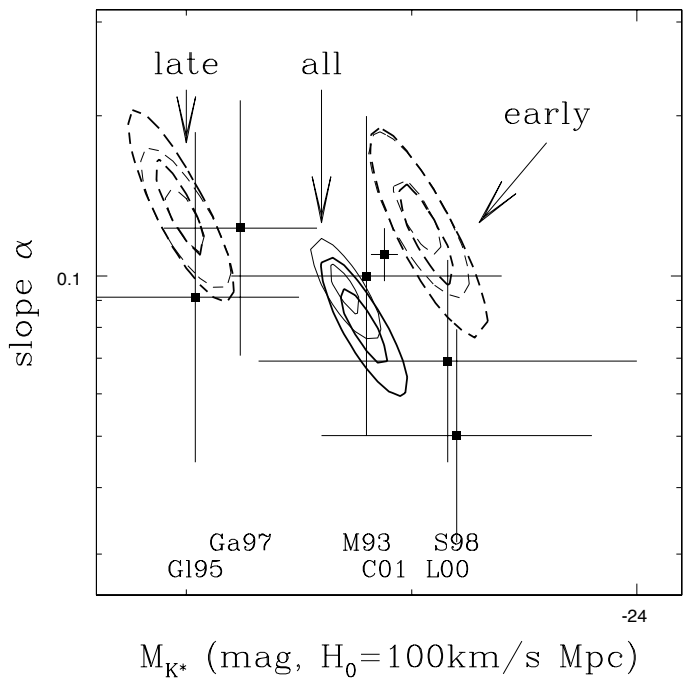


Fig. B.40. Schechter parameters α and M_* for the 2MASS luminosity functions shown in Fig. B.39. Note there is a significant correlations not only between α and M_* but also with the comoving density scale n_* that should be included in lens statistical calculations but generally are not.

efforts, none of the recent results have been presented as a survey. Surveys of all radio sources (the MIT/Greenbank survey, Burke, Lehár & Conner 1992) have the advantage that most lensed radio sources are produced by extended, steep spectrum sources (see Kochanek & Lawrence 1990), and the disadvantage that the complex intrinsic structures of extended radio sources make the follow up observations difficult. Surveys of flat spectrum radio sources (the CLASS survey, Browne et al. 2003, the PANELS survey, Winn, Hewitt & Schechter 2001) have the advantage that the follow up observations are relatively simple because most unlensed flat spectrum sources are (nearly) point sources. There are disadvantages as well – because the source structure is so simple, flat spectrum lenses tend to provide fewer constraints on mass models than steep spectrum lenses. The radio sources tend to be optically faint, making it difficult to determine their redshifts in many cases.

The second issue for any survey is to understand the method by which the sources were originally identified. For example, it is important to know

whether the flux of a lensed source in the input catalog is the total flux of all the images or only a part of the flux (e.g. the flux of the brightest image). This will have a significant effect on the statistical corrections for using a flux-limited catalog, a correction known in gravitational lensing as the "magnification bias" (see §B.6.6). All large, published surveys were essentially drawn from samples which would include the total flux of a lensed system. It is also important to know whether the survey imposed any criterion for the sources being point-like, since lensed sources are not, or any color criterion that might be violated by lensed sources with bright lens galaxies or significant extinction.

The third issue for any survey is to consider the desired selection function of the observations. This is some combination of resolution, dynamic range and field of view. These determine the range of lens separations that are detectable, the nature of any background sources, and the cost of any follow up observations. Any survey is a trade-off between completeness (what fraction of all lenses in sample that can be discovered), false positives (how many objects selected as lenses candidates that are not), and the cost of follow-up observations. The exact strategy is not critical provided it is well-understood. The primary advantages of the surveys of flat spectrum radio sources are the relatively low false positive rates and follow up costs produced by using a source population consisting almost entirely of point sources with no contaminating background population. This does not mean that the flat spectrum surveys are free of false positives – core-jet sources can initially look like asymmetric two-image lenses. On small angular scales ($\Delta\theta \lesssim 3''.0$) the quasar surveys share this advantage, but for wider separations there is contamination from binary quasars (see §B.7.2) and Galactic stars (see Kochanek 1993a).

B.6.2 The Lens Population

The probability that a source has an intervening lens requires a model for the distribution of the lens galaxies. In almost all cases these are based on the luminosity function of local galaxies combined with the assumption that the comoving density of galaxies does not evolve with redshift. Of course luminosity is not mass, so a model for converting the luminosity of a local galaxy into its deflection scale as a lens is a critical part of the process. For our purposes, the distributions of galaxies in luminosity are well-described by a Schechter (1976) function,

$$\frac{dn}{dL} = \frac{n_*}{L_*}\left(\frac{L}{L_*}\right)^{\alpha} \exp\left(-L/L_*\right). \tag{B.100}$$

The Schechter function has three parameters: a characteristic luminosity L_* (or absolute magnitude M_*), an exponent α describing the rise at low luminosity, and a comoving density scale n_*. All these parameters depend on the type of galaxy being described and the wavelength of the observations. In

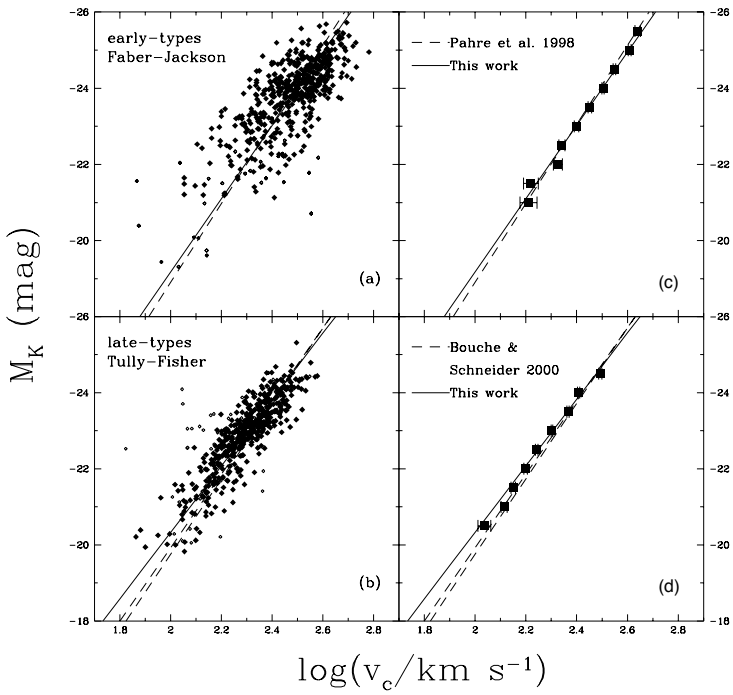


Fig. B.41. K-band kinematic relations for 2MASS galaxies. The top panels show the Faber-Jackson relation and the bottom panels show the Tully-Fisher relations for 2MASS galaxies with velocity dispersions and circular velocities drawn from the literature. The left hand panels show the individual galaxies, while the right hand galaxies show the mean relations. Note the far larger scatter of the Faber-Jackson relation compared to the Tully-Fisher relation.

general, lens calculations have divided the galaxy population into two broad classes: late-type (spiral) galaxies and early-type (E/S0) galaxies. Over the period lens statistics developed, most luminosity functions were measured in the blue, where early and late-type galaxies showed similar characteristic luminosities. The definition of a galaxy type is a slippery problem – it may be defined by the morphology of the surface brightness (the traditional method), spectral classifications (the modern method since it is easy to do in redshift surveys), colors (closely related to spectra but not identical), and stellar kinematics (ordered rotational motions versus random motions). Each approach has advantages and disadvantages, but it is important to realize that the kinematic definition is the one most closely related to gravitational lensing and the one never supplied by local surveys. Fig. B.39 shows an example of a luminosity function, in this case K-band infrared lu-

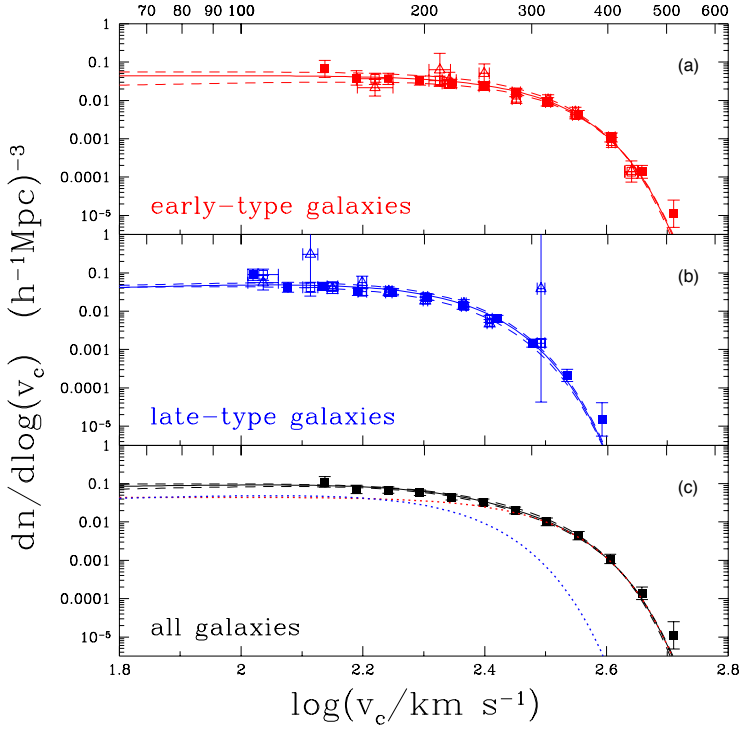


Fig. B.42. The resulting velocity functions from combining the K-band luminosity functions (Fig. B.39) and kinematic relations (Fig. B.41) for early-type (top), late-type (middle) and all (bottom) galaxies. The points show partially non-parametric estimates of the velocity function based on the binned estimates in the right hand panels of Fig. B.41 rather than power-law fits. Note that early-type galaxies dominate for high circular velocity.

minosity function by Kochanek et al. (2001, also Cole et al. 2001) where $M_{K*e} = -23.53 \pm 0.06$ mag, $n_{*e} = (0.45 \pm 0.06) \times 10^{-2} h^3$ Mpc^{-3}, and $\alpha_e = -0.87 \pm 0.09$ for galaxies which were morphologically early-type galaxies and $M_{K*l} = -22.98 \pm 0.06$ mag, $n_{*l} = (1.01 \pm 0.13) \times 10^{-2} h^3$ Mpc^{-3}, and $\alpha_l = -0.92 \pm 0.10$ for galaxies which were morphologically late-type galaxies. Early-type galaxies are less common but brighter than late-type galaxies at K-band. It is important to realize that the parameter estimates of the Schechter function are correlated, as shown in Fig. B.40, and that it is dangerous to simply extrapolate them to fainter luminosities than were actually included in the survey

However, light is not mass, and it is mass which determines lensing properties. One approach would simply be to assign a mass-to-light ratio to the galaxies and to the expected properties of the lenses. This was attempted

only in Maoz & Rix (1993) who found that for normal stellar mass to light ratios it was impossible to reproduce the data (although it is possible if you adjust the mass-to-light ratio to fit the data, also see Kochanek 1996a). Instead, most studies convert the luminosity functions dn/dL into a velocity functions dn/dv using the local kinematic properties of galaxies and then relate the stellar kinematics to the properties of the lens model. As Fig. B.41 shows (for the same K-band magnitudes of our luminosity function example), both early-type and late-type galaxies show correlations between luminosity and velocity. For late-type galaxies there is a tight correlation known as the Tully-Fisher (1977) relation between luminosity L and circular velocity v_c and for early-type galaxies there is a loose correlation known as the Faber-Jackson (1976) relation between luminosity and central velocity dispersion σ_v. Early-type galaxies do show a much tighter correlation known as the fundamental plane (Dressler et al. 1987, Djorgovski & Davis 1987) but it is a three-variable correlation between the velocity dispersion, effective radius and surface brightness (or luminosity) that we will discuss in §B.9. While there is probably some effect of the FP correlation on lens statistics, it has yet to be found. For lens calculations, the circular velocity of late-type galaxies is usually converted into an equivalent (isotropic) velocity dispersion using $v_c = \sqrt{2}\sigma_v$. We can derive the kinematic relations for the same K-band-selected galaxies used in the Kochanek et al. (2001) luminosity function, finding the Faber-Jackson relation

$$M_k - 5\log h = (-23.83 \pm 0.03) - 2.5(4.04 \pm 0.18)(\log v_c - 2.5) \qquad \text{(B.101)}$$

and the Tully-Fisher relation

$$M_k - 5\log h = (-22.92 \pm 0.02) - 2.5(3.96 \pm 0.08)(\log v_c - 2.3). \qquad \text{(B.102)}$$

These correlations, when combined with the K-band luminosity function have the advantage that the magnitude systems for the luminosity function and the kinematic relations are identical, since magnitude conversions have caused problems for a number of lens statistical studies using older photographic luminosity functions and kinematic relations. For these relations, the characteristic velocity dispersion of an L_* early-type galaxy is $\sigma_{*e} \simeq 209$ km/s while that of an L_* late-type galaxy is $\sigma_{*l} \simeq 143$ km/s. These are fairly typical values even if derived from a completely independent set of photometric data.

Both the Faber-Jackson and Tully-Fisher relations are power-law relations between luminosity and velocity, $L/L_* \propto (\sigma_v/\sigma_*)^{\gamma_{FJ}}$. This allows a simple variable transformation to convert the luminosity function into a velocity function,

$$\frac{dn}{d\sigma_v} = \frac{dn}{dL}\left|\frac{dL}{d\sigma_v}\right| = \gamma_{FJ}\frac{n_*}{\sigma_*}\left(\frac{\sigma_v}{\sigma_*}\right)^{(1+\alpha)\gamma_{FJ}-1} \exp\left(-(\sigma_v/\sigma_*)^{\gamma}_{FJ}\right). \qquad \text{(B.103)}$$

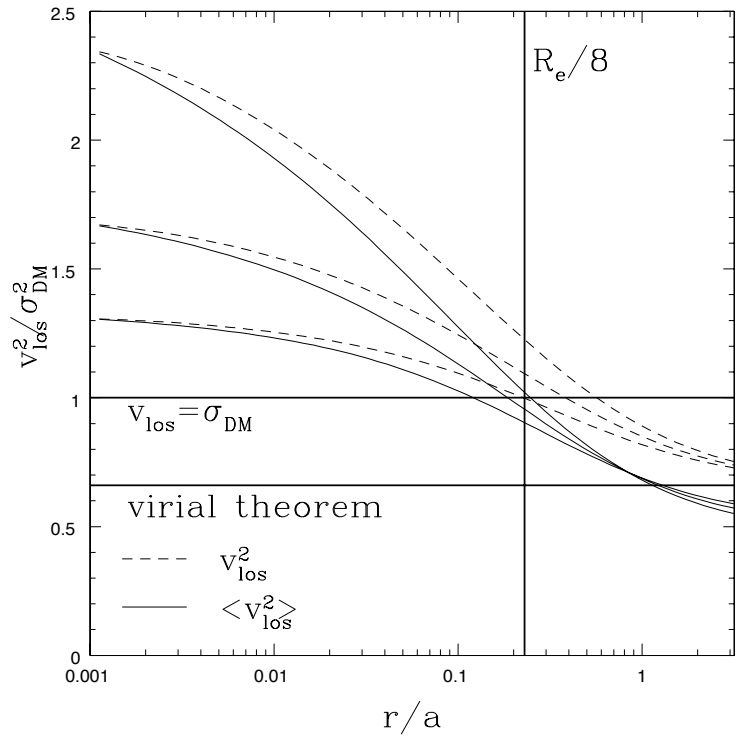


Fig. B.43. Stellar velocity dispersions v_{los} for a Hernquist distribution of stars in an isothermal halo of dispersion σ_{DM}. The solid curves show the local value v_{los} and the dashed curves show the mean interior to the radius $\langle v_{los}^2 \rangle$. Local velocity dispersions are typically measured on scales similar to $R_e/8$ where the stellar and dark matter dispersions are nearly equal rather than matching the viral theorem limit which would be reached in an infinite aperture. The upper, lower and middle curves are for stars with isotropies of $\beta = 0.2$ (somewhat radial), $\beta = 0$ (isotropic) and $\beta = -0.2$ (somewhat tangential).

There are three caveats to keep in mind about this variable change. First, we have converted to the distribution in stellar velocities, not some underlying velocity characterizing the dark matter distribution. Many early studies assumed a fixed transformation between the characteristic velocity of the stars and the lens model. In particular, Turner, Ostriker & Gott (1984) introduced the assumption $\sigma_{dark} = (3/2)^{1/2}\sigma_{stars}$ for an isothermal mass model based on the stellar dynamics (Jeans equation, Eqn. B.90 and §B.4.9) of a r^{-3} stellar density distribution in a r^{-2} isothermal mass distribution. Kochanek (1993b, 1994) showed that this oversimplified the dynamics and that if you embed a real stellar luminosity distribution in an isothermal mass distribution you actually find that the central stellar velocity dispersion is close to the veloc-

ity dispersion characterizing the dark matter halo. Fig. B.43 compares the stellar velocity dispersion to the dark matter halo dispersion for a Hernquist distribution of stars in an isothermal mass distribution. Such a normalization calculation is required for any attempt to match the observed velocity functions with a particular mass model for the lenses. Second, in an ideal world, the luminosity function and the kinematic relations should be derived from a consistent set of photometric data, while in practice they rarely are. As we will see shortly, the cross section for lensing scales roughly as σ_*^4, so small errors in estimates of the characteristic velocity have enormous impacts on the resulting cosmological results – a 5% velocity calibration error leads to a 20% error in the lens cross section. Since luminosity functions and kinematic relations are rarely derived consistently (the exception is Sheth et al. 2003), the resulting systematic errors creep into cosmological estimates. Finally, for the early-type galaxies where the Faber-Jackson kinematic relation has significant scatter, transforming the luminosity function using the mean relation as we did in Eqn. B.103 while ignoring the scatter underestimates the number of high velocity dispersion galaxies (Kochanek 1994, Sheth et al. 2003). This leads to underestimates of both the image separations and the cross sections. The fundamental lesson of all these issues is that the mass scale of the lenses should be "self-calibrated" from the observed separation distribution of the lenses rather than imposed using local observations (as we discuss below in §B.6.7).

Most lens calculations have assumed that the comoving density of the lenses does not evolve with redshift. For moderate redshift sources this only requires little evolution for $z_l < 1$ (mostly $z_l < 0.5$), but for higher redshift sources it is important to think about evolution as well. The exact degree of evolution is the subject of some debate, but a standard theoretical prediction for the change between now and redshift unity is shown in Fig. B.44 (see Mitchell et al. 2004 and references therein). Because lower mass systems merge to form higher mass systems as the universe evolves, low mass systems are expected to be more abundant at higher redshifts while higher mass systems become less abundant. For the $\sigma_v \sim \sigma_* \sim 200$ km/s galaxies which dominate lens statistics, the evolution in the number of galaxies is actually quite modest out to redshift unity, so we would expect galaxy evolution to have little effect on lens statistics. Higher mass systems evolve rapidly and are far less abundant at redshift unity, but these systems will tend to be group and cluster halos rather than galaxies and the failure of the baryons to cool in these systems is of greater importance to their lensing effects than their number evolution (see §B.7). There have been a number of studies examining lens statistics with number evolution (e.g. Mao 1991, Mao & Kochanek 1994, Rix et al. 1994) and several attempts to use the lens data to constrain the evolution (Ofek, Rix & Maoz 2003, Chae & Mao 2003, Davis, Huterer & Krauss 2003).

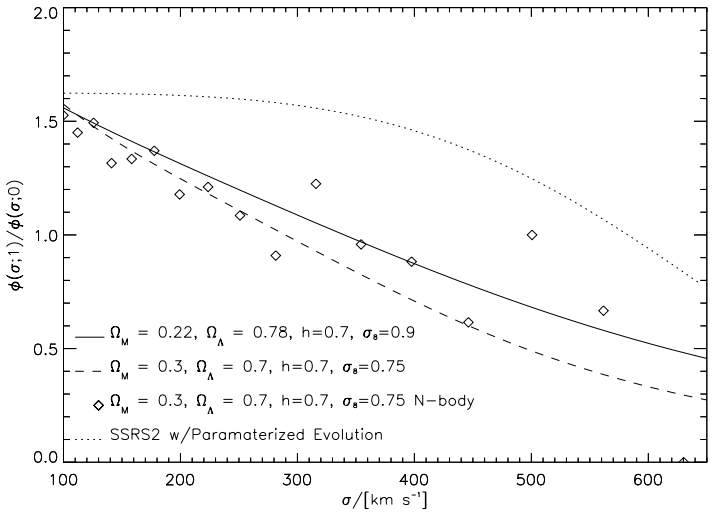


Fig. B.44. The ratio of the velocity function of halos at $z = 1$ to that at $z = 0$ from Mitchell et al. (2004). The solid curve shows the expectation for an $\Omega_\Lambda \simeq 0.78$ flat cosmological model. The points show results from an N-body simulation with $\Omega_\Lambda \simeq 0.7$ and the dashed curve shows the theoretical expectation. For comparison, the dotted curve shows the evolution model used by Chae & Mao (2003).

B.6.3 Cross Sections

The basic quantity we need for any statistical analysis is the cross section of the lens for producing the desired lensing effect (e.g. multiple images, two images, bright images...). The simplest cross section is the multiple imaging cross section of the SIS lens – the angular area on the source plane in which a source will produce two lensed images. We know from Eqns. B.21 and B.22 that the source must lie within Einstein radius b of the lens center to produce multiple images, so the cross section is simply $\sigma_{SIS} = \pi b^2$. Since the Einstein radius $b = 4\pi(\sigma_v/c)^2 D_{ls}/D_s$ depends on the velocity dispersion and redshift of the lens galaxy, we will need a model for the distribution of lenses in redshift and velocity dispersion to estimate the optical depth for lensing. If we are normalizing directly to stellar dynamical measurements of lenses, then we will also need a dynamical model (e.g. the Jeans equations of §B.4.9) to relate the observed stellar velocity dispersions to the characteristic dark matter velocity dispersion σ_v appearing as a parameter of the SIS model. We can also compute cross sections for obtaining different image morphologies. For example, in Eqn. B.32 we calculated the caustic boundaries for the four-image region of an SIS in an external shear γ. If we integrate to find the area

inside the caustic we obtain the four-image cross section

$$\sigma_4 = \frac{3\pi}{2}\frac{\gamma^2 b^2}{1-\gamma^2}, \tag{B.104}$$

while (provided $|\gamma| < 1/3$) the two-image cross section is $\sigma_2 = \sigma_{SIS} - \sigma_4 \simeq \sigma_{SIS}$. If the shear is larger, then the tips of astroid caustic extend beyond the radial (pseudo-)caustic and the lens has regions producing two images, three images in the disk geometry (Fig. B.18), and four images with no simple expression for the cross sections. There are no analytic results for the singular isothermal ellipsoid (Eqn. B.37 with $s = 0$), but we can power expand the cross section as a series in the ellipticity to find at lowest order that

$$\sigma_4 = \frac{\pi}{6} b^2 \epsilon^2 \tag{B.105}$$

for a lens with axis ratio $q = 1-\epsilon$, while the total cross section is $\sigma_{SIS} = \pi b^2$ (e.g. Kochanek 1996b, Finch et al. 2002). As a general rule, a lens of ellipticity ϵ is roughly equivalent to a spherical lens in an external shear of $\gamma \simeq \epsilon/3$. According to the cross sections, the fraction of four-image lenses should be of order $\sigma_4/\sigma_{SIS} \sim \gamma^2 \sim (\epsilon/3)^2 \sim 0.01$ rather than the observed 30%. Most of this difference is a consequence of the different magnification biases of the two image multiplicities.

There is an important subtlety when studying lens statistics with models covering a range of axis ratios, namely that the definition of the critical radius b in (say) the SIE model (Eqn. B.37) depends on the axis ratio and exactly what quantity you are holding fixed in your calculation (see Keeton, Kochanek & Seljak 1997, Keeton & Kochanek 1998, Rusin & Tegmark 2001, Chae 2003). For example, if we compare a singular isothermal sphere to a face on Mestel disk with the same equatorial circular velocity, the Einstein radius of the disk is $2/\pi$ smaller than the isothermal sphere because for the same circular velocity a disk requires less mass than a sphere. Since we usually count galaxies locally and translate these counts into a dynamical variable, this means that lens models covering a range of ellipticities must be normalized in terms of the same dynamical variables as were used to count the galaxies.

Much early effort focused of the effects of adding a finite core radius to these standard models (e.g. Blandford & Kochanek 1987b, Kochanek & Blandford 1987, Kovner 1987a, Hinshaw & Krauss 1987, Krauss & White 1992, Wallington & Narayan 1993, Kochanek 1996a). The core radius s leads to an evolution of the caustic structures (see Part 1, Blandford & Narayan 1986) with the ratio between the core radius and the critical radius s/b. Strong lenses with $s/b \ll 1$ act like singular models. Weak, or marginal, lenses with $s/b \sim 1$ have significantly reduced cross sections but higher average magnifications such that the rising magnification bias roughly balances the diminishing cross section to create a weaker than expected effect of core radii

on the probability of finding a lens (see Kochanek 1996a). As the evidence that lenses are effectively singular has mounted, interest in these models has waned, and we will not discuss them further here. There is some interest in these models as a probe of large separation lenses due to groups and clusters where a finite core radius is replaced by effects of the shallow $\rho \propto r^{-1}$ NFW density cusp, and we will consider this problem in §B.7 where we discuss large separation lenses.

B.6.4 Optical Depth

The optical depth associated with a cross section is the fraction of the sky in which you can place a source and see the effect. This simply requires adding up the contributions from all the lens galaxies between the observer and the redshift of the source. For the SIS lens we simply need to know the comoving density of lenses per unit dark matter velocity dispersion $dn/d\sigma$ (which may be a function of redshift)

$$\tau_{SIS} = \int_0^{z_s} \frac{dV}{dz_l} dz_l \int_0^{\infty} \frac{dn}{d\sigma_v} \frac{\sigma_{SIS}}{4\pi} d\sigma_v \tag{B.106}$$

where dV/dz_l is the comoving volume element per unit redshift (e.g. Turner, Ostriker & Gott 1984). For a flat cosmology, which we adopt from here on, the comoving volume element is simply $dV = 4\pi D_d^2 dD_d$ where D_d is the comoving distance to the lens redshift (Eqn. B.2). As with most lens calculations, this means that the expression simplifies if expressed in terms of the comoving angular diameter distances,

$$\tau_{SIS} = \int_0^{D_s} dD_d D_d^2 \left(\frac{D_{ds}}{D_s}\right)^2 \int_0^{\infty} \frac{dn}{d\sigma_v} 16\pi^2 \left(\frac{\sigma_v}{c}\right)^4 \tag{B.107}$$

(Gott, Park & Lee 1989, Fukugita, Futamase & Kasai 1990). If the comoving density of the lenses does not depend on redshift, the integrals separate to give

$$\tau_{SIS} = \frac{8\pi^2}{15} D_s^3 \int_0^{\infty} d\sigma_v \frac{dn}{d\sigma_v} \left(\frac{\sigma_v}{c}\right)^4 \tag{B.108}$$

(Fukugita & Turner 1991). If we now assume that the galaxies can be described by the combination of Schechter luminosity functions and kinematic relations described in §B.6.2, then we can do the remaining integral to find that

$$\tau_{SIS} = \frac{8\pi^2}{15} n_* \left(\frac{\sigma_*}{c}\right)^4 D_s^3 \Gamma[1+\alpha+6/\gamma] = \frac{1}{30} \tau_* r_H^{-3} D_s^3 \Gamma[1+\alpha+6/\gamma] \tag{B.109}$$

where $\Gamma[x]$ is a Gamma function, $r_H = c/H_0$ is the Hubble radius and the optical depth scale is

$$\tau_* = 16\pi^3 n_* r_H^3 \left(\frac{\sigma_*}{c}\right)^4 = 0.026 \left(\frac{n_*}{10^{-2} h^3 \mathrm{Mpc}^{-3}}\right) \left(\frac{\sigma_*}{200 \mathrm{km\ s}^{-1}}\right)^4 . \tag{B.110}$$

Thus, lens statistics are essentially a volume test of the cosmology (the D_s^3), predicated on knowing the comoving density of the lenses (n_*) and their average mass (σ_*). The result does not depend on the Hubble constant – all determinations of n_* scale with the Hubble constant such that $n_* D_s^3$ is independent of H_0.

Two other distributions, those in image separation and in lens redshift at fixed image separation, are easily calculated for the SIS model and useful if numerical for any other lens. The SIS image separation is $\Delta\theta = 8\pi(\sigma_v/c)^2 D_{ds}/D_s$, so

$$\frac{d\tau_{SIS}}{d\Delta\theta} = \tfrac{1}{2} D_s^3 \hat{\Delta\theta}^2 \left(\Gamma\left[1+\alpha-2/\gamma_{FJ},\xi\right]\right. \tag{B.111}$$
$$\left. -2\hat{\Delta\theta}\Gamma\left[1+\alpha-4/\gamma_{FJ},\xi\right] + \hat{\Delta\theta}^2\Gamma\left[1+\alpha-6/\gamma_{FJ},\xi\right]\right)$$

where $\xi = (\Delta\theta/\Delta\theta_*)^{\gamma_{FJ}/2}$ and

$$\Delta\theta_* = 8\pi\left(\frac{\sigma_*}{c}\right)^2 = 2\overset{''}{.}3\left(\frac{\sigma_*}{200\mathrm{km\ s}^{-1}}\right)^2 \tag{B.112}$$

is the maximum separation produced by an L_* galaxy. The mean image separation,

$$\langle\Delta\theta\rangle = \frac{\Delta\theta_*}{2}\frac{\Gamma[1+\alpha+8/\gamma]}{\Gamma[1+\alpha+6/\gamma]^{1/2}}, \tag{B.113}$$

depends only on the properties of the lens galaxy and not on cosmology. If the cosmological model is not flat, a very weak dependence on cosmology is introduced (Kochanek 1993c). For a known separation $\Delta\theta$, the probability distribution for the lens redshift becomes

$$\frac{dP}{dz_l} \propto \frac{D_d^2}{D_s}\frac{dD_d}{dz_l}\exp\left[-\left(\frac{\Delta\theta}{\Delta\theta_*}\frac{D_s}{D_{ds}}\right)^{1/2}\right] \tag{B.114}$$

(we present the result only for Schechter function $\alpha = -1$ and Faber-Jackson $\gamma_{FJ} = 4$). The location of the exponential cut off introduced by the luminosity function has a strong cosmological dependence, so the presence or absence of lens galaxies at higher redshifts dominates the cosmological limits. The structure of this function is quite different from the total optical depth, which in a flat cosmology is a slowly varying function with a mean lens distance equal to one-half the distance to the source. The mean redshift changes with cosmology because of the changes in the distance-redshift relations, but the effect is not as dramatic as the redshift distributions for lenses of known separation.

We end this section by discussing the Keeton (2002) "heresy". Keeton (2002) pointed out that if you used a luminosity function derived at intermediate redshift rather than locally, then the cosmological sensitivity of the optical depth effectively vanishes when the median redshift of the lenses matches

the median redshift of the galaxies used to derive the luminosity function. The following simple thought experiment shows that this is true at one level. Suppose there was only one kind of galaxy and we make a redshift survey and count all the galaxies in a thin shell at redshift z, finding N galaxies between z and $z+\Delta z$. The implied comoving density of the galaxies, $n = N/(\Delta z dV/dz)$, depends on the cosmological model with the same volume factor appearing in the optical depth calculation (Eqn. B.106). To the extent that the redshift ranges and weightings of the galaxy survey and a lens survey are similar, the cosmological sensitivity of the optical depth vanishes because the volume factor cancels and the optical depth depends only on the number of observed galaxies N. This does not occur when we use a local luminosity function because changes in cosmology have no effect on the local volume element. The problem with the Keeton (2002) argument is that it basically says that if we could use galaxy number counts to determine the cosmological model then we would not need lensing to do so because the two are redundant. To continue our thought experiment, we also have local estimates n_{local} for the density of galaxies, and as we vary the cosmology we would find that n and n_{local} agree only for a limited range of cosmological models and this would restore the cosmological sensitivity. The problem is that the comparison of near and distant measurements of the numbers of galaxies is tricky because it depends on correctly matching the galaxies in the presence of galaxy evolution and selection effects – in essence, you cannot use this argument to eliminate the cosmological sensitivity of lens surveys unless you think you understand galaxy evolution so well that you can use galaxy number counts to determine the cosmological model, a program of research that has basically been abandoned.

B.6.5 Spiral Galaxy Lenses

Discussions of lens statistics, or even lenses in general, focus on early-type galaxies (E/S0). The reason is that spiral lenses are relatively rare. The only morphologically obvious spirals are B0218+357 (Sc, York et al. 2004), B1600+434 (S0/Sa, Jaunsen & Hjorth 1997), PKS1830–211 (Sb/Sc, Winn et al. 2002b), PMNJ2004–1349 (Sb/Sc, Winn, Hall & Schechter 2003), and Q2237–0305 (Sa, Huchra et al. 1985). Other small separation systems may well be spiral galaxies, but we do not have direct evidence from imaging. There are studies of individual spiral lenses or the statistics of spiral lenses by Maller, Flores & Primack (1997), Keeton & Kochanek (1998), Koopmans et al. (1998), Maller et al. (2000), Trott & Webster (2002), and Winn, Hall & Schechter (2003).

The reason lens samples are dominated by early-type galaxies is that the early-type galaxies are more massive even if slightly less numerous (e.g. Fukugita & Turner 1991, see §B.6.2). The relative numbers of early-type and late-type lenses should be the ratio of their optical depths, $(n_l/n_e)(\sigma_l/\sigma_e)^4$, based on the comoving densities and characteristic velocity dispersions of

the early and late-type galaxies. For example, in the Kochanek et al. (2001) K-band luminosity function $n_l/n_e \simeq 2.2$ while the ratio of the characteristic velocity dispersions is $\sigma_{*l}/\sigma_{*e} = 0.68$ giving an expected fraction of 32% spiral. This is modestly higher than the values using other luminosity functions (usually closer to 20%) or the observed fraction. Because the typical separation of the spiral lenses will also be smaller by a factor of $(\sigma_{*l}/\sigma_{*e})^2 = 0.46$, they will be much harder to resolve given the finite resolution of lens surveys. Thus, survey selections functions discriminate more strongly against late-type lenses than against early-type lenses. The higher prevalence of dust in late-type lenses adds a further bias against them in optical surveys.

B.6.6 Magnification Bias

The optical depth calculation suggests that the likelihood of finding that a $z_s \simeq 2$ quasar is lensed is very small ($\tau \sim 10^{-4}$) , while observational surveys of bright quasars typically find that of order 1% of bright quasars are lensed. The origin of the discrepancy is the effect known as "magnification bias" (Turner 1980), which is really the correction needed to account for the selection of survey targets from flux limited samples. Multiple imaging always magnifies the source, so lensed sources are brighter than the population from which they are drawn. For example, the mean magnification of all multiply imaged systems is simply the area over which we observe the lensed images divided by the area inside the caustic producing multiple images because the magnification is the Jacobean relating area on the image and source planes, $d^2\boldsymbol{\beta} = |\mu|^{-1} d^2\boldsymbol{\theta}$. For example, an SIS lens with Einstein radius b produces multiple images over a region of radius b on the source plane (i.e. the cross section is πb^2), and these images are observed over a region of radius $2b$ on the image plane, so the mean multiple-image magnification is $\langle\mu\rangle = (4\pi b^2)/(\pi b^2) = 4$.

Since fainter sources are almost always more numerous than brighter sources, magnification bias almost always increases your chances of finding a lens. The simplest example is to imagine a lens which always produces the same magnification μ applied to a population with number counts $N(F)$ with flux F. The number counts of the lensed population are then $N_{lens}(F) = \tau\mu^{-1}N(F/\mu)$, so the fraction lensed objects (at flux F) is larger than the number expected from the optical depth if fainter objects are more numerous than the magnification times the density of brighter objects. Where did the extra factor of magnification come from? It has to be there to conserve the total number of sources or equivalently the area on the source and lens planes – you can always check your expression for the magnification bias by computing the number counts of lenses and checking to make sure that the total number of lenses equals the total number of sources if the optical depth is unity.

Real lenses do not produce unique magnifications, so it is necessary to work out the magnification probability distribution $P(>\mu)$ (the probability

of a magnification larger than μ) or its differential $dP/d\mu$ and then convolve it with the source counts. Equivalently we can define a magnification dependent cross section, $d\sigma/d\mu = \sigma dP/d\mu$ where σ is the total cross section. We can do this easily only for the SIS lens, where a source at β produces two images with a total magnification of $\mu = 2/\beta$ with $\mu > 2$ in the multiple image region (Eqns B.21, B.22), to find that $P(> \mu) = (2/\mu)^2$ and $dP/d\mu = 8/\mu^3$. The structure at low magnification depends on the lens model, but all sensible lens models have $P(> \mu) \propto \mu^{-2}$ at high magnification because this is generic to the statistics of fold caustics (Part 1, Blandford & Narayan 1986).

Usually people have defined a magnification bias factor $B(F)$ for sources of flux F so that the probability $p(F)$ of finding a lens with flux F is related to the optical depth by $p(F) = \tau B(F)$. The magnification bias factor is

$$B(F) = N(F)^{-1} \int \frac{d\mu}{\mu} \frac{dP}{d\mu} N\left(\frac{F}{\mu}\right) \tag{B.115}$$

for a source with flux F, or

$$B(m) = N(m)^{-1} \int d\mu \frac{dP}{d\mu} N\left(m + 2.5 \log \mu\right) \tag{B.116}$$

for a source of magnitude m. Note the vanishing of the extra $1/\mu$ factor when using logarithmic number counts $N(m)$ for the sources rather than the flux counts $N(F)$. Most standard models have magnification probability distributions similar to the SIS model, with $P(> \mu) \simeq (\mu_0/\mu)^2$ for $\mu > \mu_0$, in which case the magnification bias factor for sources with power law number counts $N(F) = dN/dF \propto F^{-\alpha}$ is

$$B(F) = \frac{2\mu_0^{\alpha-1}}{3-\alpha} \tag{B.117}$$

provided the number counts are sufficiently shallow ($\alpha < 3$). For number counts as a function of magnitude $N(m) = dN/dm \propto 10^{am}$ (where $a = 0.4(\alpha - 1)$) the bias factor is

$$B(F) = \frac{2\mu_0^{2.5a}}{2.5a - 2}. \tag{B.118}$$

The steeper the number counts and the brighter the source is relative to any break between a steep slope and a shallow slope, the greater the magnification bias. For radio sources a simple power law model suffices, with $\alpha \simeq 2.07 \pm 0.11$ for the CLASS survey (Rusin & Tegmark 2001), leading to a magnification bias factor of $B \simeq 5$. For quasars, however, the bright quasars have number counts steeper than this critical slope, so the location of the break from the steep slope of the bright quasars to the shallower slope for fainter quasars near $B \sim 19$ mag is critical to determining the magnification bias. Fig. B.45 shows an example of a typical quasar number counts distribution as compared to

several (old) models for the distribution of lensed quasars. The changes in the magnification bias with magnitude are visible as the varying ratio between the lensed and unlensed counts, with a much smaller ratio for bright quasars (high magnification bias) than for faint quasars (low magnification bias) and a smooth shift between the two limits as you approach the break in the slope of the counts at $B \sim 19$ mag.

For optically-selected lenses, magnification bias is "undone" by extinction in the lens galaxy because extinction provides an effect that makes lensed quasars dimmer than their unlensed counterparts. Since the quasar samples were typically selected at blue wavelengths, the rest wavelength corresponding to the quasar selection band at the redshift of the lens galaxy where it encounters the dust is similar to the U-band. If we use a standard color excess $E(B-V)$ for the amount of dust, then the images become fainter by of order $A_U E(B-V)$ magnitudes where $A_U \simeq 4.9$. Thus, if lenses had an average extinction of only $E(B-V) \simeq 0.05$ mag, the net magnification of the lensed images would be reduced by about 25%. If all lenses had the same demagnification factor $f < 1$ then the modifications to the magnification bias would be straight forward. For power-law number counts $N(F) \propto F^{-\alpha}$, the magnification bias is reduced by the factor f^α and a $E(B-V) = 0.05$ extinction leads to a 50% reduction in the magnification bias for objects with a slope $\alpha \simeq 2$ (faint quasars) and to still larger reductions for bright quasars. Some examples of the changes with the addition of a simple mean extinction are shown in the right panel of Fig. B.45, although the levels of extinction shown there are larger than observed in typical lenses as we discuss in §B.9.1. Comparisons between the statistics of optically-selected and radio-selected samples can be used to estimate the magnitude of the correction. The only such comparison found estimated extinctions consistent with the direct measurements of §B.9.1 (Falco, Kochanek & Muñoz 1998). However, the ISM of real lenses is presumably far more complicated, with a distribution of extinctions and different extinctions for different images which may be a function of orientation and impact parameter relative to the lens galaxy, for which we have no good theoretical model.

The flux of the lens galaxy also can modify the magnification bias for faint quasars, although the actual sense of the effect is complex. The left panel Fig. B.45 shows the effect of dropping lenses in which the lens galaxy represents some fraction of the total flux of the lensed images. The correction is unimportant for bright quasars because lens galaxies with $B < 19$ mag are rare. In this picture, the flux from the lens galaxy leads to the loss of lenses because the added flux from the lens galaxy makes the colors of faint lensed quasars differ from those of unlensed quasars so they are never selected as quasars to begin with. Alternatively, if one need not worry about color contamination, then the lens galaxy increases the magnification bias by supplying extra flux that makes lensed quasars brighter.

Any other selection effect, such as the dynamic range allowed for flux ratios between images as a function of their separation will also have an effect on the magnification bias. Exactly how the effect enters depends on the particular class of images being considered. For example, in the SIS lens (or more generally for two-image lenses), a limitation on the detectable flux ratio $0 < f_{min} < 1$ sets a minimum detectable magnification $\mu_{min} = 2(1 + f_{min})/(1 - f_{min}) > \mu_0 = 2$. Since most lens samples have significant magnification bias, which means that most lenses are significantly magnified, such flux limits have only modest effects. The other limit, which cannot be captured in the SIS model, is that almost all bright images are merging pairs on folds (or triplets on cusps) so the image separation decreases as the magnification increases. The contrast between the merging images and any other images also increases with increasing magnification – combined with limits on the detectability of images, these lead to selection effects against highly magnified images. This is also usually a modest effect – while magnification bias is important, the statistics are dominated by modestly magnified systems rather than very highly magnified images. In fact, there are have been few attempts at complete studies of the complicated interactions between finding quasars, finding lenses, selection effects and magnification bias. There is an early general study by Kochanek (1991b) and a detailed practical application of many of these issues to the SDSS survey by Pindor et al. (2003). Unfortunately, Pindor et al. (2003) seem to arrive at a completeness estimate from their selection model that is too high given the number of lenses they found in practice. Some of this may be due to underestimating the luminosity of lens galaxies, the effects of the lens galaxy or extinction on the selection of quasars or the treatment of extended, multicomponent lenses compared to normal quasars in the photometric pipeline. These difficulties, as well as the larger size of the present radio-selected lens samples, are the reason that almost all recent statistical studies have focused exclusively on radio lenses.

The standard magnification bias expressions (Eqns. B.115 and B.116) are not always appropriate. They are correct for the statistics of lenses selected from source populations for which the total flux of the source (including all images of a lensed source) is defining F (or m). This is true of most existing surveys – for example the CLASS radio survey sources were originally selected from single dish observations with very poor resolution compared to typical image separations (see Browne et al. 2003). If, however, the separation of the images is large compared to the resolution of the observations and the fluxes of the images are considered separately, then the bias must be computed in terms of the bright image used to select sources to search for additional images. This typically reduces the bias. More subtle effects can also appear. For example, the SDSS survey selects quasar candidates based on the best fit point-source magnitudes, which will tend to be an underestimate of the flux of a resolved lens. Hence the magnification bias for lenses found in the SDSS survey will be less than in the standard theory. Samples selected based

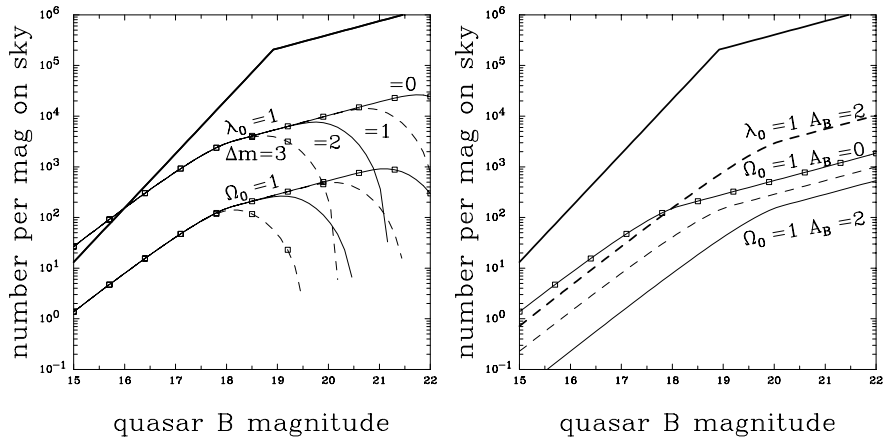


Fig. B.45. Examples of selection effects on optically selected lens samples. The heavy solid curves in the two panels shows a model for the magnitude distribution of optically-selected quasars. The light curves labeled $\Omega_M = 1$ and $\lambda_0 = 1$ show the distribution of lensed quasars for flat cosmologies that are either pure matter or pure cosmological constant. The change in the ratio between the lensed curves and the unlensed curves illustrates the higher magnification bias for bright quasars where the number count distribution is steeper than for faint quasars. In the left panel the truncated curves show the effect of losing the lensed systems where the lens galaxy is $\Delta m = 1$, 2 or 3 magnitudes fainter than the quasars. Once surveys are searching for lensed quasars with $B \gtrsim 20$ mag, the light from the lens galaxy becomes an increasing problem, particularly since the systems with the brightest lens galaxies will also have the largest image separations that would otherwise make them easily detected. In the left panel we illustrate the effect of adding a net extinction of $A_B = 1$ or 2 mag from dust in the lens galaxies. These correspond to larger than expected color excesses of $E(B-V) \simeq 0.2$ and 0.4 mag respectively. Note how the extinction "undoes" the magnification bias by shifting the lensed distributions to fainter magnitudes.

on more than one frequency can have more complicated magnification biases depending on the structure of the multidimensional number counts (Borgeest, von Linde & Refsdal 1991, Wyithe, Winn & Rusin 2003). The exact behavior is complex, but the magnification bias can be tremendously increased if the fluxes in the bands are completely uncorrelated or tightly but nonlinearly correlated. For example, if the luminosities in bands A and B are related by tight, nonlinear correlation of the form $L_A \propto L_B^{1/2}$, then the lensed examples of these objects will lie off the correlation. At present, there are too few deep, wide-area multiwavelength catalogs to make good use of this idea, but this is changing rapidly.

In general, the ellipticity of the lenses has little effect on the expected number of lenses, allowing the use of circular lens models for statistical stud-

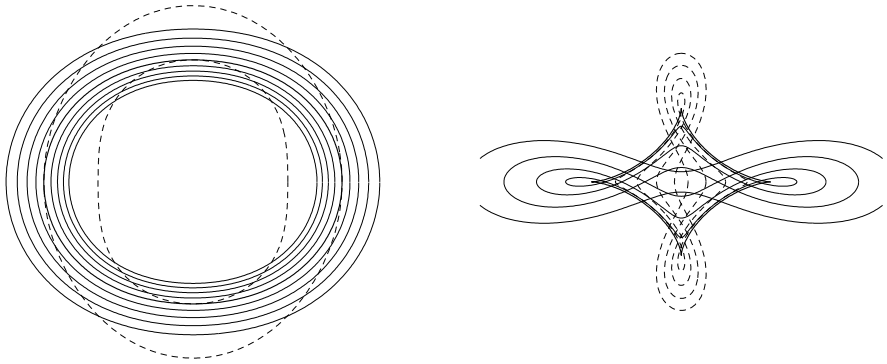


Fig. B.46. Magnification contours on the image (left) and source (right) planes for an SIS in an external shear. The heavy solid contours show the tangential critical line (left) and its corresponding caustic (right). On the image plane (left), the light curves are magnification contours. These are positive outside the critical curve and negative inside the critical curve. The images found in a four-image lens are all found in the region between the two dashed contours – when two images are merging on the critical line, the other two images lie on these curves. On the source plane the solid (dashed) curves show the projections of the positive (negative) magnification contours onto the source plane. Note that the high magnification regions are dominated by the four-image systems with the exception of the small high magnification regions found just outside the tip of each cusp.

ies that are uninterested in the morphologies of the images (e.g. Keeton, Kochanek & Seljak 1997, Rusin & Tegmark 2001, Chae 2003). However, the effects of ellipticity are trivially observable in the relative numbers of two-image and four-image lenses. We noted earlier that the expectation from the cross section is that four-image lenses should represent order $\epsilon_\Psi^2 \sim \gamma^2 \sim 0.01$ of lenses where ϵ_Ψ is the ellipticity of the lens potential. Yet in §B.2 we saw that four-image lenses represent roughly one third of the observed population. The high abundance of four-image lenses is a consequence of the different magnification biases of the two-image multiplicities – the four-image lenses are more highly magnified than the two-image lenses so they have a larger magnification bias factor.

Fig. B.46 shows the image magnification contours for an SIS lens in an external shear on both the image and source planes. The highly magnified regions are confined to lie near the critical line. If we Taylor expand the inverse magnification radially, then $\mu^{-1} = \Delta x |d\mu^{-1}/dx|$ where Δx is the distance from the critical line, so the magnification drops inversely with the distance from the critical line. If we Taylor expand the lens equations, then we find

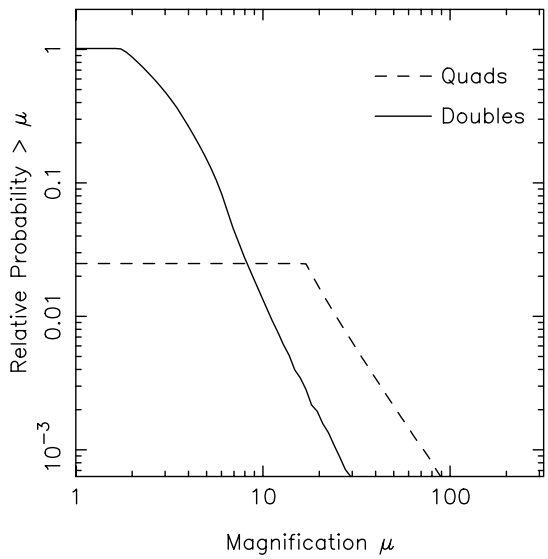


Fig. B.47. The integral magnification probability distributions for a singular isothermal ellipsoid with an axis ratio of $q = 0.7$ normalized by the total cross section for finding two images. Note that the total four-image cross section is only of order $\epsilon_\Psi^2 \sim (\epsilon/3)^2 \sim 0.01$ of the total, but that the minimum magnification for the four-image systems ($\mu_{min} \sim 1/\epsilon \sim 10$) is much larger than that for the two-image systems ($\mu_{min} \sim 2$ just as for an SIS). The entire four-image probability distribution is well approximated by the $P(> \mu) \propto \mu^{-2}$ power law expected for fold caustics, while the two-image probability distribution is steeper since highly magnified images can only be created by the cusps. Figure courtesy of D. Rusin.

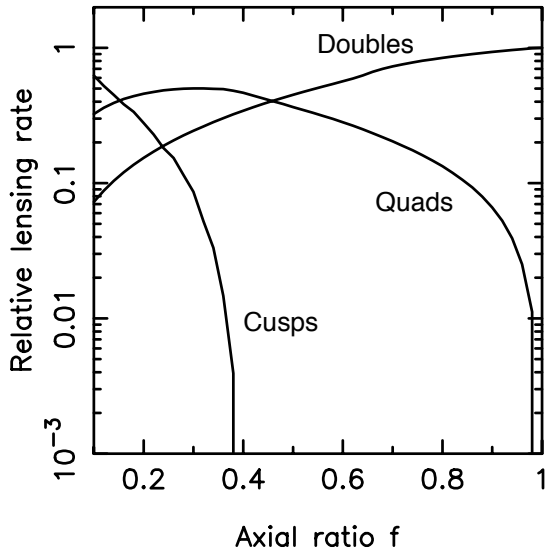


Fig. B.48. The expected number of two-image, four-image and three-image (disk or cusp) lenses as a function of axis ratio f for the CLASS sample. From Rusin & Tegmark (2001).

that the change in source plane coordinates is related to the change in image plane coordinates by $\Delta\beta = \mu^{-1}\Delta x \propto \mu^{-2}$. Thus, if L is the length of the astroid curve, the probability of a magnification larger than μ scales as $P(>\mu) \propto \mu^{-2}L/|d\mu^{-1}/dx|$. This applies only to the four image region, because the only way to get a high magnification in the two-image region is for the source to lie just outside the tip of a cusp. The algebra is overly complex to present, but the generic result is that the region producing magnification μ extends μ^{-2} from the cusp tip but has a width that scales as $\mu^{-1/2}$, leading to an overall scaling that the asymptotic cross section declines as $P(>\mu) \propto \mu^{-7/2}$ rather than $P(>\mu) \propto \mu^{-2}$. This can all be done formally (see Blandford & Narayan 1986) so that asymptotic cross sections can be derived for any model (e.g. Kochanek & Blandford 1987, Finch et al. 2002), but a reasonable approximation for the four-image region is to compute the magnification, μ_0, for the cruciform lens formed when the source is directly behind the lens and then use the estimate that $P(>\mu) = (\mu_0/\mu)^2$. Unfortunately, such simple estimates are not feasible for the two-image region. These distributions are relatively easy to compute numerically, as in the example shown in Fig. B.47.

Because the minimum magnification of a four-image lens increases $\mu_0 \propto \gamma^{-1}$ even as the cross section decreases as $\sigma_4 \propto \gamma^2$, the expected number of four-image lenses in a sample varies much more slowly with ellipticity than expected from the cross section. The product $\sigma_4 B(F) \propto \gamma^2\mu_0^{\alpha-1}$, of the four-image cross section, σ_4, and the magnification bias, $B(F)$, scales as $\gamma^{3-\alpha} \propto \gamma$ for the CLASS survey ($\alpha \simeq 2$), which is a much more gentle dependence on ellipticity than the quadratic variation expected from the cross section. There is a limit, however, to the fraction of four-image lenses. If the potential becomes too flat, the astroid caustic extends outside the radial caustic (Fig. B.18), to produce three-image systems in the "disk" geometry rather than additional four-image lenses. In the limit that the axis ratio goes to zero (the lens becomes a line), only the disk geometry is produced. The existence of a maximum four-image lens fraction, and its location at an axis ratio inconsistent with the observed axis ratios of the dominant early-type lenses has made it difficult to explain the observed fraction of four image lenses (King & Browne 1996, Kochanek 1996b, Keeton, Kochanek & Seljak 1997, Keeton & Kochanek 1998,Rusin & Tegmark 2001). Recently, Cohn & Kochanek (2001) argued that satellite galaxies of the lenses provide the explanation by somewhat boosting the fraction of four-image lenses while at the same time explaining the existence of the more complex lenses like B1359+154 (Myers et al. 1999, Rusin et al. 2001) and PMNJ0134–0931 (Winn et al. 2002c, Keeton & Winn 2003) formed by having multiple lens galaxies with more complex caustic structures. It is not, however, clear in the existing data that four-image systems are more likely to have satellites to the lens galaxy than two-image systems as one would expect for this explanation.

Gravitational lenses can produce highly magnified images without multiple images only if they are highly elliptical or have a low central density.

The SIS lens has a single-image magnification probability distribution of $\tau dP/d\mu = 2\pi b^2/(\mu-1)^3$ with $\mu < 2$ compared to $\tau dP/d\mu = 2\pi b^2/\mu^3$ with $\mu \geq 2$ for the multiply imaged region, so single images are never magnified by more than a factor of 2. For galaxies, where we always expect high central densities, the only way to get highly magnified single images is when the astroid caustic extends outside the radial caustic (Fig. B.18). A source just outside an exposed cusp tip can be highly magnified with a magnification probability distribution $dP/d\mu \propto \mu^{-7/2}$. Such single image magnifications have recently been a concern for the luminosity function of high redshift quasars (e.g. Wyithe 2004, Keeton, Kuhlen & Haiman 2004) and will be the high magnification tail of any magnification perturbations to supernova fluxes (e.g. Dalal et al. 2003). As a general rule for galaxies, the probability of a single image being magnified by more than a factor of two is comparable to the probability of being multiply imaged.

B.6.7 Cosmology With Lens Statistics

The statistics of lenses, in the sense of the number of lenses expected in a sample of sources as a function of cosmology, is a volume test of the cosmological model because the optical depth (at least for flat cosmologies) is proportional to D_s^3. However, the number of lenses also depends on the comoving density and mass of the lenses (n_*, σ_* and α in the simple SIS model). While n_* could plausibly be estimated locally, the σ_*^4 dependence on the mass scale makes it very difficult to use local estimates of galaxy kinematics or masses to normalize the optical depth. The key step to eliminating this problem is to note that there is an intimate relation between the cross section, the observed image separations and the mass scale. While this will hold for any mass model, the SIS model is the only simple analytic example. The mean image separation for the lenses should be independent of the cosmological model for flat cosmologies (and only weakly dependent on it otherwise). Thus, in any lens sample you can eliminate the dependence on the mass scale by replacing it with the observed mean image separation, $\tau_{SIS} \propto n_* \langle \Delta\theta \rangle^2 D_s^3$. Full calculations must include corrections for angular selection effects. Most odd results in lens cosmology arise in calculations that ignore the close coupling between the image separations and the cross section.

In practice, real calculations are based on variations of the maximum likelihood method introduced by Kochanek (1993b, 1996a). For each lens i you compute the probability p_i that it is lensed including magnification bias and selection effects. The likelihood of the observations is then

$$\ln L_0 = \sum_{lenses} \ln p_i + \sum_{unlensed} \ln(1-p_i) \simeq \sum_{lenses} \ln p_i - \sum_{unlensed} p_i \quad \text{(B.119)}$$

where $\ln(1-p_i) \simeq -p_i$ provided $p_i \ll 1$. This simply encodes the likelihood of finding the observed number of lenses given the individual probabilities that

the objects are lensed. Without further information, this likelihood could determine the limits on the cosmological model only to the extent we had accurate prior estimates for n_* and σ_*.

If we add, however, a term for the probability that each detected lens has its observed separation (Eqn. B.112 plus any selection effects)

$$\ln L = \ln L_0 + \sum_{lensed} \ln \left(\frac{p_i(\Delta\theta_i)}{p_i} \right), \tag{B.120}$$

then the lens sample itself can normalize the typical mass scale of the lenses (Kochanek 1993b). This has two advantages. First, it eliminates any systematic problems arising from the dynamical normalization of the lens model and its relation to the luminosity function. Second, it forces the cosmological estimates from the lenses to be consistent with the observed image separations – it makes no sense to produce cosmological limits that imply image separations inconsistent with the observations. In theory the precision exceeds that of any local calibration very rapidly. The fractional spread of the separations about the mean is ~ 0.7, so the fractional uncertainty in the mean separation scales as $0.7/N^{1/2}$ for a sample of N lenses. Since the cross section goes as the square of the mean separation, the uncertainty in the mean cross section $1.4/N^{1/2}$ exceeds any plausible accuracy of a local normalization for σ_* (10% in σ_*, or 20% in $\langle\theta\rangle \propto \sigma_*^2$, or 40% in $\tau \propto \sigma_*^4$) with only $N \simeq 10$ lenses.

Any other measurable property of the lenses can be added to the likelihood, but the only other term that has been seriously investigated is the probability of the observed lens redshift given the image separations and the source redshift (Kochanek 1992a, 1996a, Helbig & Kayser 1996, Ofek, Rix & Maoz 2003). In general, cosmologies with a large cosmological constant predict significantly higher lens redshifts than those without, and in theory this is a very powerful test because of the exponential cutoff in Eqn. B.114. The biggest problem in actually using the redshift test, in fact so big that it probably cannot be used at present, is the high incompleteness of the lens redshift measurements (§B.2). There will be a general tendency, even at fixed separation, for the redshifts of the higher redshift lens galaxies to be the ones that are unmeasured. Complete samples could be defined for a separation range, usually by excluding small separation systems, but a complete analysis needs to include the effects of groups and cluster boosting image separations beyond the splitting produced by an isolated galaxy. For example, how do we include Q0957+561 with its separation of $6\farcs2$ that is largely due to the lens galaxy but has significant contributions from the surrounding cluster?

B.6.8 The Current State

Recent analyses of lens statistics have focused exclusively on the CLASS flat spectrum radio survey (Browne et al. 2003). Chae et al. (2002), Chae (2003) and Mitchell et al. (2004) focus on estimating the cosmological model and find

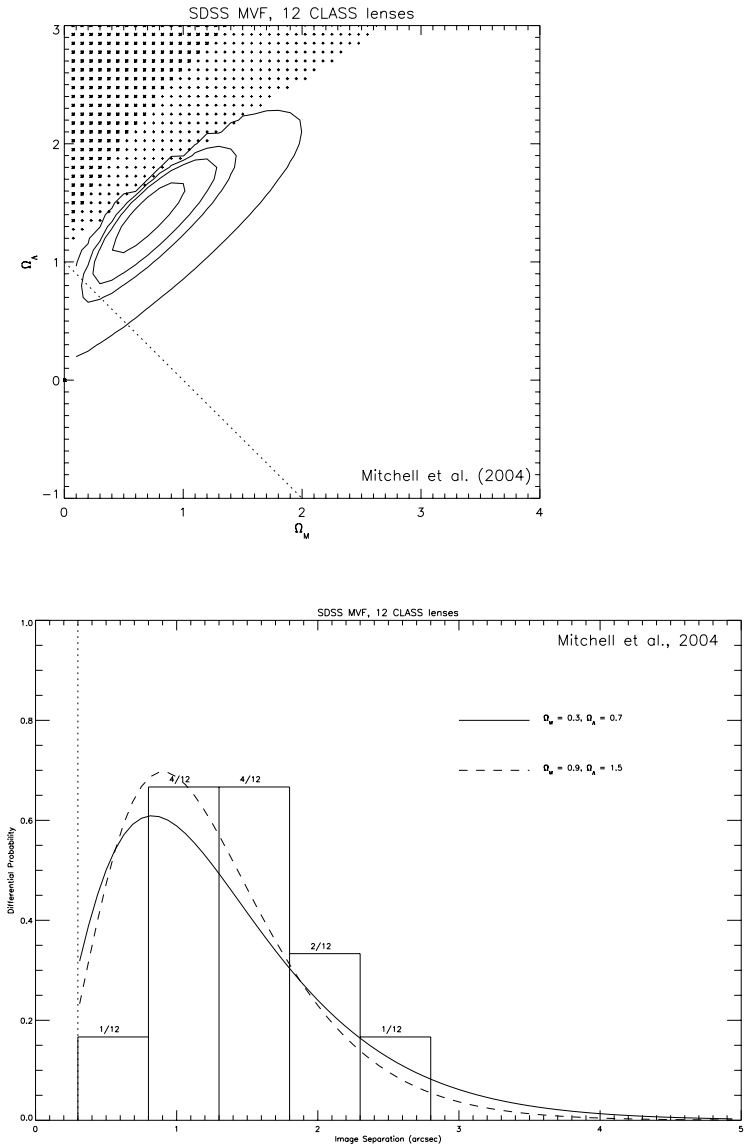


Fig. B.49. (Top) Likelihood functions for the cosmological model from Mitchell et al. (2004) using the velocity function of galaxies measured from the SDSS survey and a sample of 12 CLASS lenses. The contours show the 68, 90, 95 and 99% confidence intervals on the cosmological model. In the shaded regions the cosmological distances either become imaginary or there is no big bang. (Bottom) The histogram shows the separation distribution of the 12 CLASS lenses used in the analysis and the curve shows the distribution predicted by the maximum likelihood model including selection effects.

results in general agreement with estimates from Type Ia supernovae (e.g. Riess et al. 2004). The general approach of both groups is to use variants of the maximum likelihood methods described above in §B.6.7. Chae (2003) uses an obsolete estimate of the galaxy luminosity function combined with a Faber-Jackson relation and the variable transformation of Eqn. B.103 but normalized the velocity scale using the observed distribution of lens separations. Mitchell et al. (2004) use the true velocity dispersion function from the SDSS survey (Sheth et al. 2003) and incorporate a Press-Schechter (1974) model for the evolution of the velocity function. Chae (2003) used ellipsoidal galaxies, although this has little cosmological effect, while Mitchell et al. (2004) considered only SIS models. Fig. B.49 shows the cosmological limits from Mitchell et al. (2004), which are typical of the recent results . There are also attempts to use lens statistics to constrain dark energy (e.g. Chae et al. 2004, Kuhlen, Keeton & Madau 2004), but far larger, well-defined samples are needed before the resulting constraints will become interesting.

Chae & Mao (2003), Davis, Huterer & Krauss (2003) and Ofek, Rix & Maoz (2003) focused on galaxy properties and evolution in a fixed, concordance cosmology rather than on determining the cosmological models. Mitchell et al. (2004) compared models where the lenses evolved following the predictions of CDM models in comparison to non-evolving models. Because lens statistical estimates are unlikely to complete with other means of estimating the cosmological models, these are more promising applications of gravitational lens statistics for the future. Attempts to estimate the evolution of the lens population usually allow the n_* and σ_* parameters of the velocity function (Eqn. B.103) to evolve as power laws with redshift. Mitchell et al. (2004, Fig. B.44) point out that CDM halo models make specific predictions for the evolution of the velocity function that have a different structure from simple power laws in redshift, but with the present data the differences are probably unimportant. All these evolution studies came to the conclusion that the number density of the $\sigma_v \sim \sigma_*$ galaxies which dominate lens statistics has changed little ($\lesssim \pm 50\%$) between the present day and redshift unity.

I have three concerns about these analyses and their focus on the "complete" CLASS lens samples. First, a basic problem with the CLASS survey is that we lack direct measurements of the redshift distribution of the source population forming the lenses (e.g Marlow et al. 2000, Muñoz et al. 2003). In particular, Muñoz et al. (2003) note that the radio source population is changing radically from nearly all quasars to mostly galaxies as you approach the fluxes of the CLASS source population. This makes it dangerous to extrapolate the source population redshifts from the brighter radio fluxes where the redshift samples are nearly complete to the fainter samples where they are not. The second problem is that no study has a satisfactory treatment of the lenses with satellites or associated with clusters. All the analyses use isolated lens models and then either include lenses with satellites but ignore

the satellites or drop lenses with satellites and ignore the fact that they have been dropped. The analysis by Cohn & Kochanek (2004) of lens statistics with satellites shows that neither approach is satisfactory – dropping the satellites biases the results to underestimate cross sections while including them does the reverse. Cohn & Kochanek (2004) concluded that including they systems with satellites probably has fewer biases than dropping them. A similar problem probably arises from the effects of the group halos to which many of the lenses belong (e.g. Keeton et al. 2000, Fassnacht & Lubin 2002). My third concern is that the separations of the radio lenses seen to be systematically smaller than the optically selected lenses even though the optical HST Snapshot survey (Maoz et al. 1993) had the greatest sensitivity to small separation systems. It is possible that this is simply due to selection effects in the optical samples, but I have seen no convincing scenario for producing such a selection effect. We see no clear correlation of extinction with image separation (see §B.9.1), emission from the lens galaxy is less important for small separation systems than for large separation systems, and the selection function due to the resolution of the observations is fairly simple to model.

On the other hand, the various lens samples may all consistent. One way to compare the different data sets is to non-parametrically construct the velocity function from the observed image separations of the samples. To do this we assume an SIS lens model for the conversion from image separations to circular velocities, and then adopt the standard non-parametric methods used to construct luminosity functions from redshift surveys to construct the velocity function from the image separations (Kochanek 2003c). The results for the flat-spectrum lens surveys (CLASS, JVAS, PANELS), all radio surveys and all radio surveys plus the quasar lenses are shown in Fig. B.50. We normalized the estimates to the density at $v_c = 300$ km/s to eliminate any dependence on the cosmological model. The lens data can estimate the velocity function from roughly $v_c \sim 100$ km/s to 500 km/s. At lower velocities the finite resolution of the observations makes the uncertainties in the density explode, and at higher velocities the surveys have not searched large enough angular regions around the lens galaxies. The shape of the velocity function is consistent with local estimates (Fig. B.42) except in the highest circular velocity bin where we begin to see the contribution from clusters we will consider in §B.7. Fig. B.50 also makes it clear why constraints on the evolution of the lenses are so weak – evolution estimates basically try to compare the low-redshift separation distribution to the high redshift separation distribution, and we simply do not have large enough lens samples to begin subdividing them in redshift (to say nothing of dealing with unmeasured redshifts) and still have small statistical uncertainties.

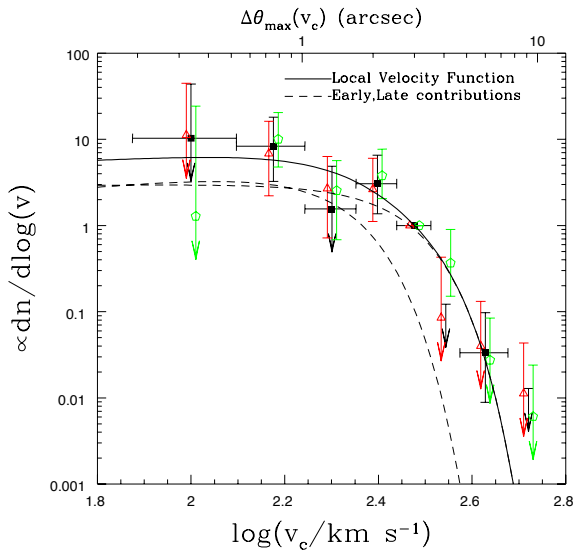


Fig. B.50. Non-parametric reconstructions of the velocity function from the observed separations of gravitational lenses assuming an SIS lens model. The velocity functions are all normalized to the bin centered at 300 km/s. The filled squares use only the lenses in the flat spectrum radio surveys, the triangles use all radio-selected lenses and the pentagons include all radio lenses and all quasar lenses. The horizontal error bars on the filled squares show the bin widths. The triangles and pentagons are horizontally offset from the squares to make them more visible. The curves show the velocity function estimated from the 2MASS sample from Fig. B.42. The horizontal scale at the top of the figure shows the maximum separation produced by a lens of the corresponding circular velocity. The mean separation produced by such a lens will be one-half the maximum.

B.7 What Happened to The Cluster Lenses?

One would think from the number of conference proceeding covers featuring HST images of cluster arcs that these are by far the most common type of lens. In fact, this is an optical delusion created by the ease of finding the rich clusters even though they are exponentially rare. The most common kind of lens is the one produced by a typical massive galaxy – as we saw in in Fig. B.50. For a comparison, Fig. B.51 shows several estimates of the velocity function based on standard CDM mass functions and halo models (from Kochanek & White 2001 and Kochanek 2003c, using the Sheth & Tormen 1999 mass function combined with the NFW halo model from §B.4.1). We see for high masses or circular velocities that the predicted distribution of halos agrees

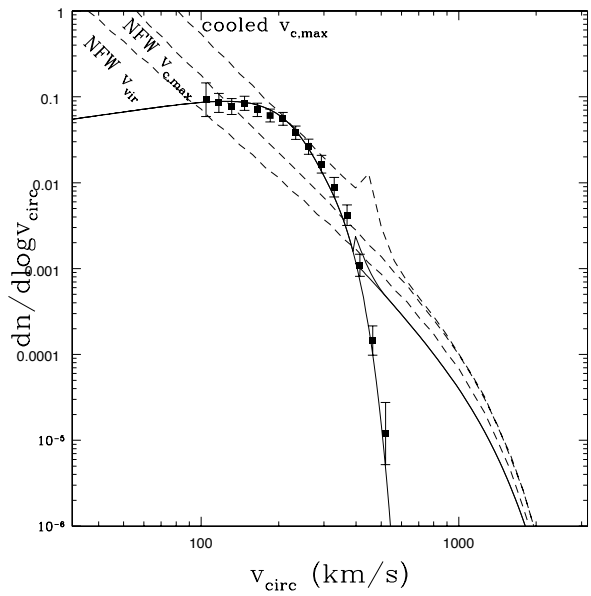


Fig. B.51. The expected circular velocity function $dn/d\log v_c$ of CDM halos The lowest dashed curve labeled NFW v_{vir} shows the velocity function using the NFW halo virial velocity v_{vir} for the circular velocity (see §B.4.1). The middle dashed curve labeled NFW $v_{c,max}$ shows the velocity function if the peak circular velocity of the halo is used rather than the virial velocity. The upper dashed curve is a model in which the baryons of halos with $M \lesssim 10^{13} M_\odot$ cool, raising the central density and circular velocity. The solid curve with the points shows the estimate of the local velocity function of galaxies (Fig. B.42) and the solid curve extending to higher velocities is an estimate of the local velocity function of groups and clusters.

with the observed distribution of clusters. At the velocities typical of galaxies, the observed density of galaxies is nearly an order of magnitude higher than expected for a CDM halo mass function. At very low velocities we expect many more halos than we observe galaxies. The velocity function estimated from the observed image separations matches that of galaxies with the beginnings of a tail extending onto the distribution of clusters at the high velocity end (Fig. B.51). At low velocities the limited resolution of the present surveys means that the current lens data does not probe the low velocity end very well. In this section we discuss the difference between cluster and galaxy lenses and explain the origin of the break between galaxies and clusters. In §B.8 on CDM substructure we will discuss the divergence at low circular velocities.

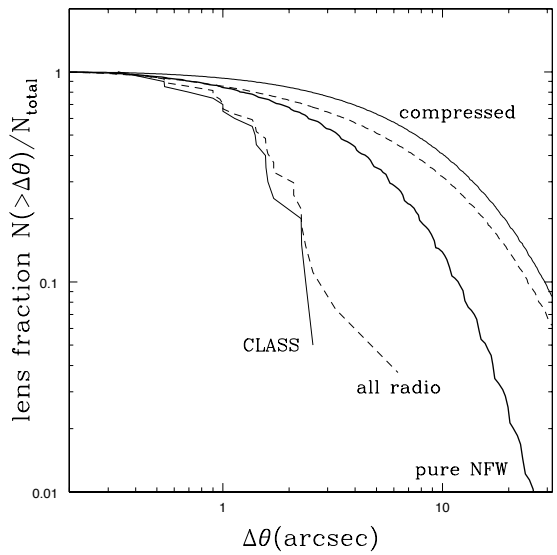


Fig. B.52. Predicted image separation distributions assuming the structure of halos does not change with halo mass. The heavy solid line shows the prediction for pure NFW models while the light solid (dashed) curves shows the predictions after 5% of the baryons have cooled into a disk (a disk plus a bulge with 10% of the baryonic mass in the bulge). The curves labeled CLASS (for the CLASS survey lenses) and all radio (for all radio selected lenses) show the observed distributions.

The standard halo mass function is roughly a power law with $dn/dM \sim M^{-1.8}$ combined with an exponential cutoff at the mass scale corresponding to the largest clusters that could have formed at any epoch (e.g. the Sheth & Tormen 1999 halo mass function). Typically these rich clusters have internal velocity dispersions above 1000 km/s and can produce image splittings of ~ 30 arcsec. If halo structure was independent of mass, then we would expect the separation distribution of gravitational lenses to show a similar structure – a power law out to the mass scale of rich clusters followed by an exponential cutoff. In Fig. B.52 we compare the observed distribution of radio lenses to that expected from the halo mass function assuming either NFW halos or NFW halos in which the baryons, representing 5% of the halo mass, have cooled and condensed into the centers of the halos (Kochanek & White 2001). We would find similar curves if we used simple SIS models rather than these more complex CDM-based models (Keeton 1998, Porciani & Madau 2000). In practice, the most complete survey for multiply imaged sources, the CLASS survey, found a largest separation of $4\farcs5$ (B2108+213) despite checking candidates out to separations of $15\farcs0$ (Phillips et al. 2001). The largest lens found in a search for multiply imaged sources has an image separation of roughly 15 arcsec (SDSS1004+4112, Inada et al. 2003). The overall separation distri-

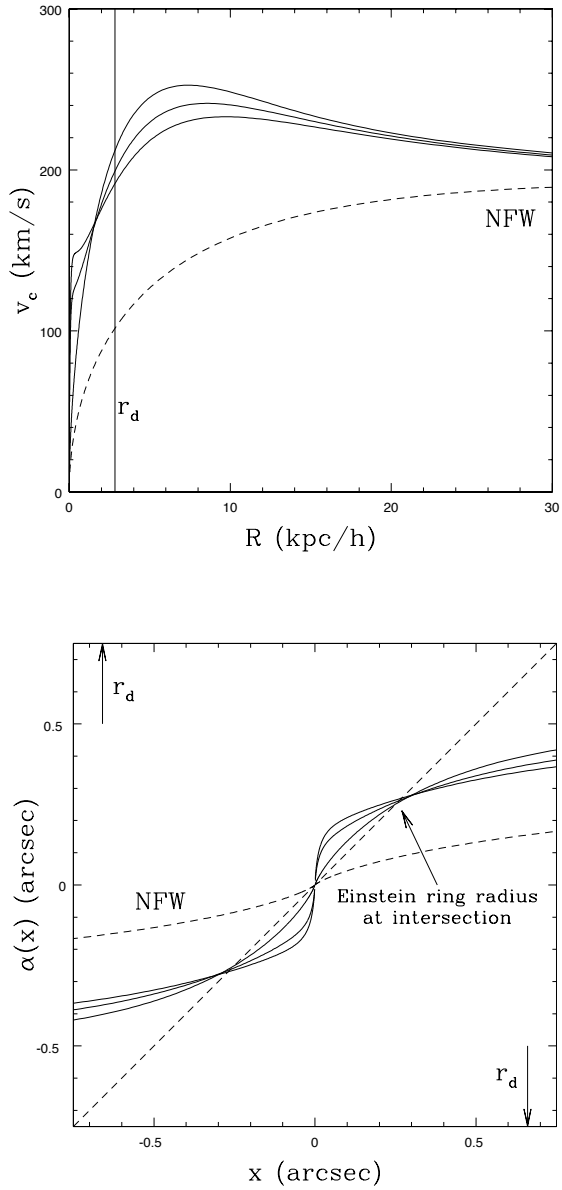
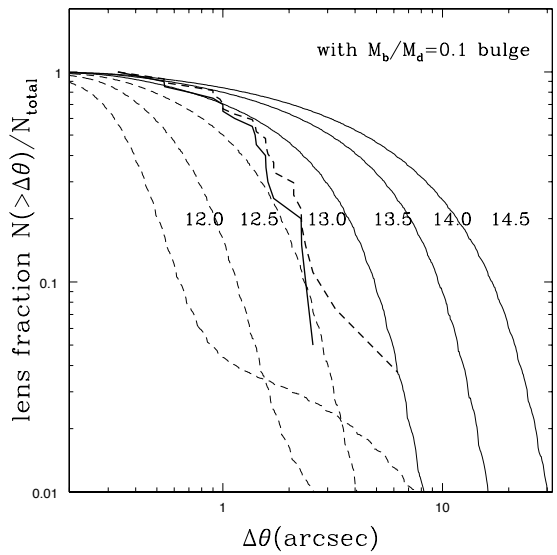


Fig. B.53. (Top) The rotation curve and (Bottom) the bending angle $\alpha(x)$ for a $10^{12} M_\odot$ halo at $z_l = 0.5$ with a concentration of $c = 8$ lensing a source at $z_s = 2.0$. The dashed curves show the results for the initial NFW halo, while the solid curves show the results after allowing 5% of the mass to cool conserving angular momentum (spin parameter $\lambda = 0.04$) and adiabatically compressing the dark matter. The three solid curves show the effect of putting 0%, 10% or 20% of the baryonic mass into a central bulge. Higher bulge masses raise the central circular velocity and steepen the central deflection profile. The final disk scale length is r_d. Compare these to the bending angles of our simple models in Figs. B.10–B.14.

bution (see Fig. B.52) has a sharp cutoff on scales of 3 arcsec corresponding to galaxies with velocity dispersions of ~ 250 km/s. The principal searches for wide separation lenses are Maoz et al. (1997), Ofek et al. (2001) and Phillips et al. (2001), although most surveys searched for image separations of at least $6\farcs0$. A large number of studies focused only on the properties of lenses produced by CDM mass functions (e.g. Narayan & White 1988, Wambsganss et al. 1995, 1998, Kochanek 1995b, Maoz et al. 1997, Flores & Primack 1996, Mortlock & Webster 2000b, Li & Ostriker 2002, Keeton & Madau 2001, Wyithe, Turner & Spergel 2001). We will not discuss these in detail because such models cannot reproduce the observed separation distributions of lenses. Most recent analyses allow for changes in the density distributions between galaxies and clusters.

Physically the important difference between galaxies and clusters is that the baryons in the galaxies have cooled and condensed into the center of the halo to form the visible galaxy. As the baryons cool, they also drag some of the dark matter inward through a process known as adiabatic compression (Blumenthal et al. 1986), although this is less important than the cooling. As we show in Fig. B.53, standard dark matter halos are terrible lenses because their central cusps ($\rho \propto r^{-\gamma}$ and $1.5 \geq \gamma \geq 1$) are too shallow. In this case, a standard NFW halo with a total mass of $10^{12} M_\odot$ and a concentration of $c = 8$ (see Eqns. B.60–B.62) at a redshift of $z_l = 0.5$ is unable to produce multiple images of a source at redshift $z_s = 2$ despite having an asymptotic circular velocity of nearly 200 km/s. If we now assume that 5% of the mass is in baryons starting with a typical halo angular momentum and then cooling into a disk of radius r_d while conserving angular momentum we see that the rotation curve becomes flatter and the galaxy is now able to produce multiple images. Putting some fraction of the mass into a still more compact, central bulge make the lens even more supercritical and the bending angle diagram begins to resemble that of an SIS lens (see Fig. B.11). Thus, the cooling of the baryons converts a sub-critical dark matter halo into one capable of producing multiple images.

The key point is that only intermediate mass halos contain baryons which have cooled. High mass halos (groups and clusters) have cooling times longer than the Hubble time so they have not had time too cool (e.g. Rees & Ostriker 1977). Most low mass halos also probably resemble dark matter halos more than galaxies with large quantities of cold baryons because they lost their baryons due to heating from the UV background during the initial period of star formation (e.g. Klypin et al. 1999 Bullock, Kravtsov & Weinberg 2000, see §B.8). Here we ignore the very low mass halos and consider only the distinction between galaxies and groups/clusters. The fundamental realization in recent studies (e.g. Porciani & Madau 2000, Kochanek & White 2001, Kuhlen, Keeton & Madau 2004, Li & Ostriker 2003) is that introducing a cooling mass scale M_c below which the baryons cool to form galaxies and above which they do not supplies the explanation for the difference be-



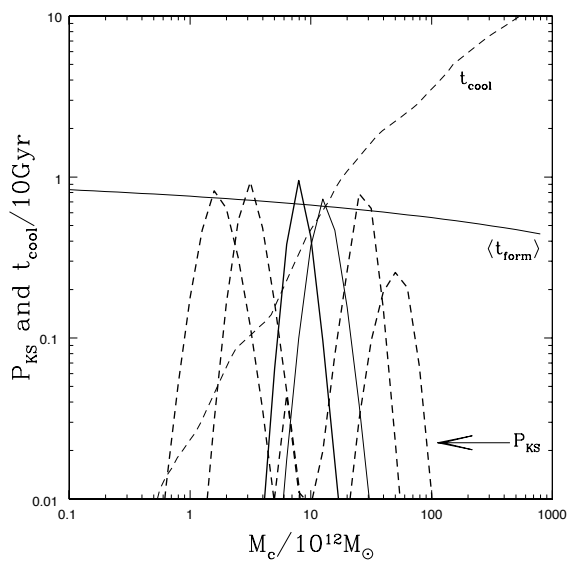


Fig. B.54. (Top) Predicted separation distributions as a function of the cooling mass scale M_c in which 5% of the mass cools with 90% of the cooled material in a disk and 10% in a bulge. The dashed curves show the distributions for $M_c = 10^{12}M_\odot$, $3 \times 10^{12}M_\odot$ and $10^{13}M_\odot$, while the solid curves show the distributions for $M_c = 3 \times 10^{13}M_\odot$, $10^{14}M_\odot$ and $3 \times 10^{14}M_\odot$. The heavy solid (dashed) curves shows the observed distribution of the CLASS (all radio-selected) lenses.

Fig. B.55. (Bottom) The Kolmogorov-Smirnov probability, P_{KS}, of fitting the observed distribution of CLASS lenses as a function of the cooling mass scale M_c. The heavy solid curves show the results when 5% of the mass cools without (with) 10% of that mass in a bulge. The heavy dashed curves show the results for models where lower (1% and 2%) or higher (10% and 20%) halo mass fractions cool, where the optimal cooling mass scale M_c decreases as the cold baryon fraction increases. For comparison, the light dashed line shows the cooling time t_{cool} in units of 10 Gyr for the radius enclosing 50% of the baryonic mass in the standard model. The light solid line shows the average formation epoch, $\langle t_{form} \rangle$, also in units of 10 Gyr.

tween the observed separation distribution of lenses and naive estimates from halo mass functions.

Once we recognize the necessity of introducing a distinction between cluster and galaxy mass halos, we can use the observed distribution of lens separations to constrain the mass scale of the break and the physics of cooling. Fig. B.54 shows the most common version of these studies, where separation distributions are computed as a function of the cooling mass scale M_c. We show the separation distributions for various cooling mass scales assuming that 5% of the mass cools into a disk plus a bulge with 10% of the baryonic mass in the bulge for all halos with $M < M_c$. If the cooling mass is either too low or too high we return to the models of Fig. B.52, while at some intermediate mass scale we get the break in the separation distribution to match the observed angular scale. For these parameters, the optimal cooling mass scale is $M_c \simeq 10^{13} M_\odot$ (Fig. B.55). This agrees reasonably well with Porciani & Madau (2000) and Kuhlen, Keeton & Madau (2004) who found a somewhat higher mass scale $M_c \simeq 3 \times 10^{13} M_\odot$ using SIS models for galaxies. Cosmological hydrodynamic simulations by Pearce et al. (1999) also found that approximately 50% of the baryons had cooled on mass scales near $10^{13} M_\odot$. Note, however, that the mass scale needed to fit the data depends on the assumed fraction of the mass in cold baryons. With fewer cold baryons a halo becomes a less efficient lens producing smaller image separations so M_c must increase to keep the break at the observed scale. If the cold baryon fraction is too low ($\lesssim 1\%$), it becomes impossible to explain the data at all. Crudely, the cooling mass scale depends exponentially on the cold baryon fraction with $\log M_c/M_\odot \simeq 13.6 - (\text{cold fraction})/0.15$.

The mass scale of the break and the cold baryon fraction are not independent parameters and should be derivable from the physics of the cooling gas. In its full details this must include not only the cooling of the gas but also reheating of the gas in galaxies due to feedback from star formation. Fig. B.55 also shows the dependence of the cooling time scale and the formation time scale for halos of mass M_c. For this model (based on the semi-analytic models of Cole et al. 2000), the cooling time becomes shorter than the age of the halo very close to the mass scale required to explain the distribution of image separations. These semi-analytic models suggest an alternate approach in where the cooling mass scale need not be added as an *ad hoc* parameter. We could instead follow the semi-analytic models and use the cooling function to determine the relative cooling rates of halos with different masses. We leave as the free parameter, the final cosmological density in cold baryons $\Omega_{b,cool} \leq \Omega_b \simeq 0.04$ (i.e. some baryons may never cool or cool and are reheated by feedback). Low $\Omega_{b,cool}$ models have difficulty cooling, making them equivalent to models with a high cooling mass scale. High $\Omega_{b,cool}$ models cool easily, making them equivalent to models with a high cooling mass scale. Models with $0.015 \lesssim \Omega_{b,cool} \lesssim 0.025$ agree with the observations. The result depends little on whether we add a bulge, fit the CLASS sample or all

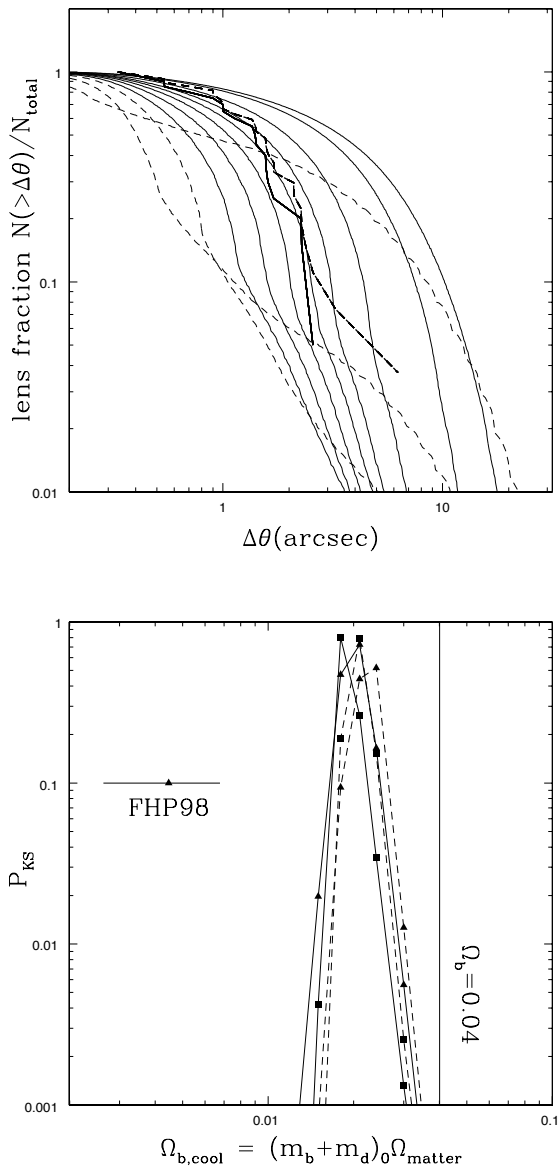


Fig. B.56. (Top) Predicted separation distributions as a function of the cosmological cold baryon density $\Omega_{b,cool}$. The dashed curves show the results for $\Omega_{b,cool} = 0.003$, 0.006 and 0.009 (right to left at large separation) and the solid curves show the results for $\Omega_{b,cool} = 0.0012$, 0.015, 0.018, 0.021, 0.024, 0.030, 0.045 and 0.060 (from left too right at large separation). The models have 10% of the cold baryons in a bulge. The heavy solid (dashed) curves show the observed distribution of CLASS (all radio) lenses.

Fig. B.57. (Bottom) The Kolmogorov-Smirnov probability, P_{KS}, of fitting the observed distribution of lenses as a function of the cold baryon density $\Omega_{b,cool}$. The squares (triangles) indicate models with no bulge (10% of the cooled material in a bulge), and the solid (dashed) lines correspond to fitting the CLASS (all radio) lenses. For comparison, the horizontal error bar is the estimate by Fukugita, Hogan & Peebles (1998) for the cold baryon (stars, remnants, cold gas) content of local galaxies. The vertical line marks the total baryon content of the concordance model.

radio lenses or adjust the cooling curve by a factor of two. Thus, the characteristic scale of the gravitational lens separation distribution is a probe of the cosmological baryon density Ω_b and the fraction of those baryons that cool in the typical massive galaxy. While it would be premature to use this as a method for determining Ω_b, it is interesting to note that our estimate is significantly below current cosmological estimates that $\Omega_b \simeq 0.04$ which would be consistent with feedback from star formation and other processes preventing all baryons from cooling, but well above the estimates of the cold baryon fraction in local galaxies ($0.0045 \lesssim \Omega_{b,cool} \lesssim 0.0068$, Fukugita, Hogan & Peebles 1998). These are also the models generating the velocity function estimate with baryonic cooling in Fig. B.51. The cooling of the baryons shifts the more numerous low velocity halos to higher circular velocities so that the models match the observed density of $\sigma_v \simeq \sigma_*$ galaxies. The models do not correctly treat the break region because they allow "over-cooled" massive groups, but then merge back onto the peak circular velocity distribution of the CDM halos at higher velocities. Since the models allow all low mass halos to cool, there is still a divergence at low circular velocities which is closely related to the problem of CDM substructure we discuss in §B.8.

B.7.1 The Effects of Halo Structure and the Power Spectrum

Estimating the structure of clusters using gravitational lensing is primarily a topic for Part 3, so we include only an abbreviated discussion of lensing by clusters here. For a fixed cosmological model, two parameters largely control the abundance of cluster lenses. First, the abundance of clusters varies nearly exponentially with the standard normalization $\sigma_8 \simeq 1$ of the power spectrum on $8h^{-1}$ Mpc scales. Second, the cross sections of the individual clusters depend strongly on the exponent of the central density cusp of the cluster. There are recent studies of these issues by Li & Ostriker (2002, 2003), Huterer & Ma (2004), Kuhlen, Keeton & Madau (2004), Oguri et al. (2004), and Oguri & Keeton (2004).

We can understand the general effects of halo structure very easily from our simple power law model in Eqn. B.9. In §B.3 we normalized the models to have the same Einstein radius, but we now want to normalize them to all have the same total mass interior to some much larger radius R_0. This is roughly what happens when we keep the virial mass and break radius of the halo constant but vary the central density exponent $\rho \propto r^{-n}$. The deflection profile becomes

$$\alpha(\theta) = \frac{b_0^2}{R_0}\left(\frac{\theta}{R_0}\right)^{2-n} \tag{B.121}$$

where $b_0 \ll R_0$ sets the mass interior to R_0 and we recover our old example if we let $b = b_0 = R_0$. The typical image separation is determined by the tangential critical line at $\theta_t = R_0(b_0/R_0)^{2/(n-1)}$, so more centrally concentrated lenses (larger n) produce larger image separations when $b_0/R_0 \ll 1$.

The radial caustic lies at $\beta_r = f(n)\theta_t$ where $f(n)$ is a not very interesting function of the index n, so the cross section for multiple imaging $\sigma \propto \beta_r^2 \propto R_0^2(b_0/R_0)^{4/(n-1)}$ – for an SIS profile $\sigma \propto b^4/R_0^2$, while the cross section for a Moore profile ($n = 3/2$) $\sigma \propto b^8/16R_0^6$ is significantly smaller. We cannot go to the limit of an NFW profile ($n = 1$) because our power law model has a constant surface density rather than a logarithmically divergent surface density in the limit as $n \to 1$, but we can see that as the density profile becomes shallower the multiple image cross section drops rapidly when the models have constant mass inside a radius which is much larger than their Einstein radius. As a result, the numbers of group or cluster lenses depends strongly on the central exponent of the density distribution even when the mass function of halos is fixed. Magnification bias will weaken the dependence on the density slope because the models with shallower slopes and smaller cross sections will generally have higher average magnifications. The one caveat to these calculations is that many groups or clusters will have central galaxies, and the higher surface density of the galaxy can make the central density profile effectively steeper than the CDM halo in isolation.

B.7.2 Binary Quasars

Weedman et al. (1982) reported the discovery of the third "gravitational lens", Q2345+007, a pair of $z = 2.15$ quasars separated by $7\farcs3$. The optical spectra of the two images are impressively similar (e.g. Small et al. 1997), but repeated attempts to find a lens have failed in both the optical (e.g. Pello et al. 1996) and with X-rays (Green et al. 2002). Q2345+007 is the founding member of a class of objects seen in the optical as a pair of quasars with very similar spectra, small velocity differences and separations $3\farcs0 \lesssim \Delta\theta \lesssim 15\farcs0$. The most recent compilation contained 15 examples (Mortlock, Webster & Francis 1999). The incidence of these quasar pairs in surveys is roughly 2 per 1000 LBQS quasars (see Hewett et al. 1998) and 1 per 14000 CLASS radio sources (Koopmans et al. 2000). The separations of these objects correspond to either very massive galaxies or groups/clusters. Obvious lenses on these scales, in the sense that we see the lens, are rare but have an incidence consistent with theoretical expectations (see Fig. B.52). If, however, even a small fraction of the objects like Q2345–007 are actually gravitational lenses, then dark lenses outnumber normal groups and clusters and dominate the halo population on mass scales above $M \gtrsim 10^{13} M_\odot$.

If the criterion of possessing a visible lens is dropped, so as to allow for dark lenses, proving objects are lenses becomes difficult. There are two unambiguous tests – measuring a time delay between the images, which is very difficult given the the long time delays expected for lenses with such large separations, or using deep imaging to find that the host galaxies of the quasars show the characteristic arcs or Einstein rings of lensed hosts (Figs. B.3, B.4).

The latter test is feasible with HST[7] and will be trivial with JWST. Spectral comparisons have been the main area of debate. In the optical, many of the pairs have alarmingly similar spectra if they are actually binary quasars (e.g. Q2345+007 or Q1634+267, see Small et al. 1997) – indeed, some of these dark lens candidates have more similar spectra than genuinely lensed quasars (see Mortlock, Webster & Francis 1999). The clearest examples of dark lens candidates that have to be binary quasars are the cases in which only one quasar is radio loud. These objects, such as PKS1145–071 (Djorgovski et al. 1987) or MGC2214+3550 (Muñoz et al. 1998), represent 4 of the 15 candidates. Similarly, the dramatic difference in the flux ratio between optical and X-ray wavelengths of Q2345+007 is the strongest direct argument for this object being a binary quasar (Green et al. 2002).

Two statistical arguments provide the strongest evidence that these objects must be binary quasars independent of any weighting of spectral similarities. The first argument, due to Kochanek, Muñoz & Falco (1999), is that the existence of binary quasars like MGC2214+3550 in which only one of the quasars is radio loud predicts the incidence of pairs in which both are radio quiet. We can label the quasar pairs as either O^2R^2, where both quasars are seen in the optical (O) and the radio (R), O^2R, where only one quasar is seen in the radio, or O^2 where neither quasar is seen in the radio. Lenses must be either O^2R^2 or O^2 pairs. Surveys of quasars find that only $P_R \simeq 10\%$ of quasars are radio sources with 3.6 cm fluxes above 1 mJy (e.g. Bischof & Becker 2000). If all the quasar pairs were binary quasars and the probability of being radio loud is independent of whether a quasar is in a binary, then the relative number of O^2, O^2R and O^2R^2 binaries should be 1 to $2P_R = 0.2$ to $P_R^2 = 0.01$. Given that we observed 4 O^2R binaries we should observe 20 O^2 binaries and 0.2 O^2R^2 binaries. This statistical pattern matches the data, and Kochanek, Muñoz & Falco (1999) found that the most probable solution was that all quasar pairs were binary quasars with an upper limit of only 8% (68% confidence) on the fraction that could be dark lenses. With the subsequent expansion of the quasar pair sample and the discovery of the first O^2R^2 binary (B0827+525, Koopmans et al. 2000), these limits could be improved.

The second statistical argument is that the dark lens candidates do not have the statistical properties expected for lenses. Three aspects of the quasar pairs make them unlikely to be lenses simply given the properties of gravitational lensing. First, there are no four-image dark lens candidates even though a third of the normal lenses are quads. Second, many of the dark lens candidates have very high flux ratios between the images – 4 of the 9 ambiguous quasar pairs considered by Rusin (2002) have flux ratios of greater than 10:1. Magnification bias makes such large flux ratios very improbable for

[7] We detected the host galaxies of the Q2345–007 quasars in the CASTLES H-band image. Their morphology is probably inconsistent with the lens hypothesis, but we viewed the data as too marginal to publish the result

true gravitational lenses (§B.6.6, Kochanek 1995b). Third, the suppression of central/third/odd images in the lens population is a consequence of baryonic cooling and the resulting increase of the central surface density. Standard dark matter halos with their shallow central cusps, $\rho \propto r^{-1}$, generally produce detectable third images. Since it is probably a requirement for a lens to remain dark that the baryons in the halo cannot cool (or they would form stars), you would expect the typical dark lens to resemble APM08279+5255 and have an easily detectable third image (Rusin 2002). Thus, in the context of CDM we would expect dark lenses to be standard cuspy density distributions like the NFW model (Eqn. B.60). Rusin (2002) evaluated the likelihood of the quasar pairs assuming that dark lenses have the structure of CDM halos and found that the observed flux ratios and the lack of three-image dark lenses were extremely unlikely. Only the real lens APM08279+5255 had a significant probability of being produced by a dark CDM halo, although for this case I think the exposed cusp/disk lens explanation for the morphology is more likely.

The evidence overwhelmingly favors interpreting the quasar pairs as binary quasars. However, as originally pointed out by Djorgovski (1991), the one problem with the binary hypothesis is that the incidence of the quasar pairs is two orders of magnitude above that expected from an extrapolation of the quasar-quasar correlation function on scales of Mpc. As discussed in Kochanek, Muñoz & Falco (1999) and Mortlock, Webster & Francis (1999) the incidence can be increased if the incipient merger of the two host galaxies is triggering the quasar activity. The separation distribution of the binary quasars is crudely compatible with tidally triggered activity when the merger starts followed by a coalescence of the host galaxies driven by tidal friction. Small separation binary quasars ($\Delta\theta < 3''.0$) are rare because the decay of the host galaxy orbits accelerates as their separation diminishes. Well-measured angular distributions of binary quasars, potentially obtainable from SDSS, might allow detailed explorations of the triggering and merging physics.

B.8 The Role of Substructure

Simulations of CDM halos predicted many more small satellites than were actually observed in the Milky Way (e.g. Kauffmann et al. 1993, Moore et al. 1999, Klypin et al. 1999). Crudely 5-10% of the mass was left in satellites with perhaps 1-2% at the projected separations of 1–2R_e where we see most lensed images (e.g. Zentner & Bullock 2003, Mao et al. 2004). This is far larger than the observed fraction of 0.01–0.1% in observed satellites (e.g. Chiba 2002). Solutions were proposed in three broad classes: hide the satellites by preventing star formation so they are present but dark (e.g. Klypin et al. 1999, Bullock et al. 2000), destroy them using self-interacting dark matter (e.g. Spergel & Steinhardt 2000), or avoid forming them by changing the power spectrum to something similar to warm dark matter with signifi-

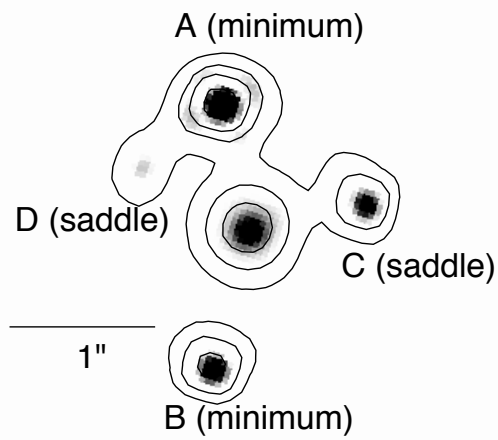


Fig. B.58. The most spectacular example of an anomalous flux ratio, SDSS0924+0219 (Inada et al. 2003). In this CASTLES infrared HST image, the D image should be comparable in brightness to the A image, but is actually an order of magnitude dimmer. The A and B images are minima, while C and D are saddle points. The contours are spaced by factors of two from the peak of the A image. The lens galaxy is seen at the center. At present we do not know whether the suppression of the saddle point in this lens is due to microlensing or substructure. If it is microlensing, ongoing monitoring programs should see it return to its expected flux within approximately 10 years.

cantly less power on the relevant mass scales (e.g. Bode et al. 2001). These hypotheses left the major observational challenge of distinguishing dark satellites from non-existent ones. This became known as the CDM substructure problem.

It was well known in the lensing community that the fluxes of lensed images were usually poorly fit by lens models. There was a long litany of reasons for ignoring them arising from possible systematic errors which can corrupt image fluxes. Differential effects between the images from the interstellar medium of the lens can corrupt the fluxes (dust in the optical/IR, scatter broadening in the radio, see §B.9.1). Time delays combined with source variability can corrupt any single-epoch measurement. Microlensing by the stars in the lens galaxy can modify the fluxes of any sufficiently compact component of the source (at a minimum the quasar accretion disk, see Part 4). The most peculiar problem was the of anomalous flux ratios in radio lenses. Radio sources are essentially unaffected by the ISM of the lens galaxy in low resolution observations that minimize the effects of scatter broadening (VLA rather than VLBI), true absorption appears to be rare, radio sources

generally show little variability even when monitored, and most of the flux should come from regions too large to be affected by microlensing. Yet in B1422+231, for example, the three cusp images violated the cusp relation for their fluxes (that the sum of the signed magnifications of the three images should be zero, see Metcalf & Zhao 2002, Keeton, Gaudi & Petters 2003, or Schneider, Ehlers & Falco 1992).[8]

It is easier to outline the problem of anomalous flux ratios near a fold caustic (such as images A and D in SDSS0924+0219, see Fig. B.58), than a cusp caustic. Near a fold, the lens equations can be reduced to a one-dimensional model with

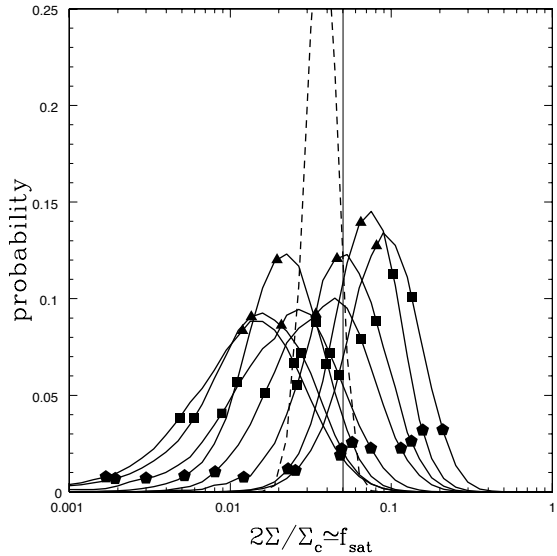
$$\beta = \theta\,(1 - \Psi'') - \frac{1}{2}\Psi'''\theta^2 \to -\frac{1}{2}\Psi'''\theta^2 \tag{B.122}$$

and inverse magnification

$$\mu^{-1} = (1 - \Psi'') - \Psi'''\theta \to -\Psi'''\theta \tag{B.123}$$

where we choose our coordinates such that there is a critical line at $\theta = 0$ (i.e. $1 - \Psi'' = 0$) and the primes denote derivatives of the potential. These equations are easily solved to find that you have images at $\theta_\pm = \pm(-2\beta/\Psi''')^{1/2}$ if the argument of the square root is positive and no solutions otherwise – as you cross the fold caustic ($\beta = 0$) two images are created or destroyed on the critical line at $\theta = 0$. Their inverse magnifications of $\mu_\pm^{-1} = \mp(-2\beta\Psi''')^{1/2}$ are equal in magnitude but reversed in sign. Hence, if the assumptions of the Taylor expansion hold, the images merging at a fold should have identical fluxes. Either by guessing or by tedious algebra you can determine that the fractional correction to the magnification from the next order term is of order $\theta_\pm\Psi^{(4)}/\Psi'''$. For any reasonable central potential where the images are at radius θ_0 from the lens center, the fractional correction will be of order $\theta_\pm/\theta_0 \sim 0.1$ for the typical pair of anomalous images. Hence, using gravity to produce the anomalous flux ratios requires terms in the potential with a length scale comparable to the separation of the images to significantly violate the rule that they should have similar fluxes. Mao & Schneider (1998) pointed out that a very simple way of achieving this was to put a satellite near the images, and they found that this could explain the anomaly in B1422+231. Metcalf & Madau (2001, also see Bradac et al. 2002 for images of the magnification patterns expected from a CDM halo) put these two pieces together, pointing out that if normal satellite galaxies were too rare to make anomalous flux ratios common, the missing CDM substructure was not. They predicted that in CDM, anomalous flux ratios should be common.

[8] In specific models there can also be global invariants relating image positions and magnifications (e.g. Witt & Mao 2000, Hunter & Evans 2001, Evans & Hunter 2002). These results are usually for simple softened power law models using either ellipsoidal potentials or an external shear rather than ellipsoidal cuspy density distributions with an external shear, so their applicability to the observed lenses is unclear.



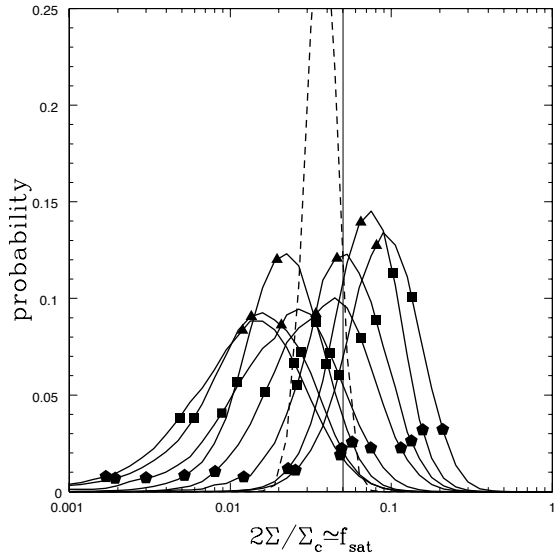


Fig. B.59. (Top) A Monte Carlo test for estimating substructure surface densities. The heavy curves show the estimated probability distribution for the substructure surface density fraction in a sample of 7 four-image lenses in which the input fraction was 5% (marked by the vertical line). The points on the curve show the median, 1σ and 2σ confidence limits. The output distributions are consistent with the true input fraction. The dashed line shows how the accuracy would improve given a sample of 56 lenses (i.e. multiplying the 8 trials of 7 images each).

Fig. B.60. (Bottom) The same method applied to the real data. The three distributions show the effects of changing assumptions on the actual flux measurement errors – the greater the measurement uncertainties the less substructure surface density is required to explain the flux ratio anomalies. The middle case (10%) is probably slightly too conservative (20% is ridiculously conservative and 5% is probably too optimistic).

If we add a population of satellites with surface density $\kappa_{sat} = \Sigma_{sat}/\Sigma$ near the images we can estimate the nature of the perturbations. If we model them as pseudo-Jaffe potentials with critical radius b and break radius[9] $a = (bb_0)^{1/2}$, then the satellites produce a deflection perturbation of order

$$\langle \delta\theta^2 \rangle^{1/2} \sim 10^{-3} b_0 \left(\frac{10\Sigma_{sat}}{\Sigma_c} \right)^{1/2} \left(\frac{10^3 b}{b_0} \right)^{3/4} . \tag{B.124}$$

Only massive satellites will be able to produce deflection perturbations large enough to be detected given typical astrometric errors. Because the astrometric constraints for lenses are so accurate, generally better than $0\farcs005$, satellites with deflection scales larger than $b \gtrsim 10^{-2} b_0$ will usually have observable effects on model fits and must be included in the basic lens model. The shear perturbation

$$\langle \delta\gamma^2 \rangle^{1/2} \sim 0.1 \left(\frac{10\Sigma_{sat}}{\Sigma_c} \right)^{1/2} \left(\frac{10^3 b}{b_0} \right)^{1/4} \left(\frac{\ln \Lambda}{10} \right)^{1/2} , \tag{B.125}$$

where $\ln \Lambda = \ln(a/s)$ is a Coulomb logarithm required to make the integral converge at small separations, is significantly larger. The effects of substructure gain on those from the primary lens as we move to quantities requiring more derivatives of the potential because the substructure has less mass but shorter length scales. For example most astronomical objects have masses and sizes that scale with internal velocity σ_v as $M \propto \sigma_v^4$ and $R \propto \sigma_v^2$. So time delays, which depend on the (two-dimensional) potential $\Psi \propto M \propto \sigma_v^4$, will be completely unaffected by substructure. Deflections, which require one spatial derivative of the potential, $\alpha \propto \Psi/R \propto \sigma_v^2$, are affected only be the more massive substructres. Magnifications, which require two spatial derivatives of the potential, $\kappa \sim \gamma \sim \Psi/R^2 \propto \sigma_v^0$, are affected equally by all mass scales provided the Einstein radius of the object is larger than the characteristic size of the source. Substructure will also affect brighter images more than fainter images because the magnifications of the brighter images are more unstable to small perturbations. Recall that the magnification $\mu = (\lambda_+ \lambda_-)^{-1}$ where one of the eigenvalues $\lambda_\pm = 1 - \kappa \pm \gamma$, usually λ_-, is small for a highly magnified image. If we now add a shear perturbation $\delta\gamma$, the perturbation to the magnification is of order $\delta\gamma/\lambda_-$ so you have a bigger fractional perturbation to the magnification for the same shear perturbation if the image is more highly magnified. The last important effect from substructure, for which I know of no simple, qualitative explanation, is that substructure discriminates between saddle points and minima when it is a small fraction of the total surface density (Schechter & Wambsganss 2002, Keeton 2003b). In this regime, the magnification distributions for the saddle points develop an extended tail toward demagnification that is not present for the minima.

[9] This is the tidal truncation radius for an SIS of critical radius b orbiting in an SIS of critical radius $b_0 > b$. The total satellite mass is $\simeq \pi a b \Sigma_c$.

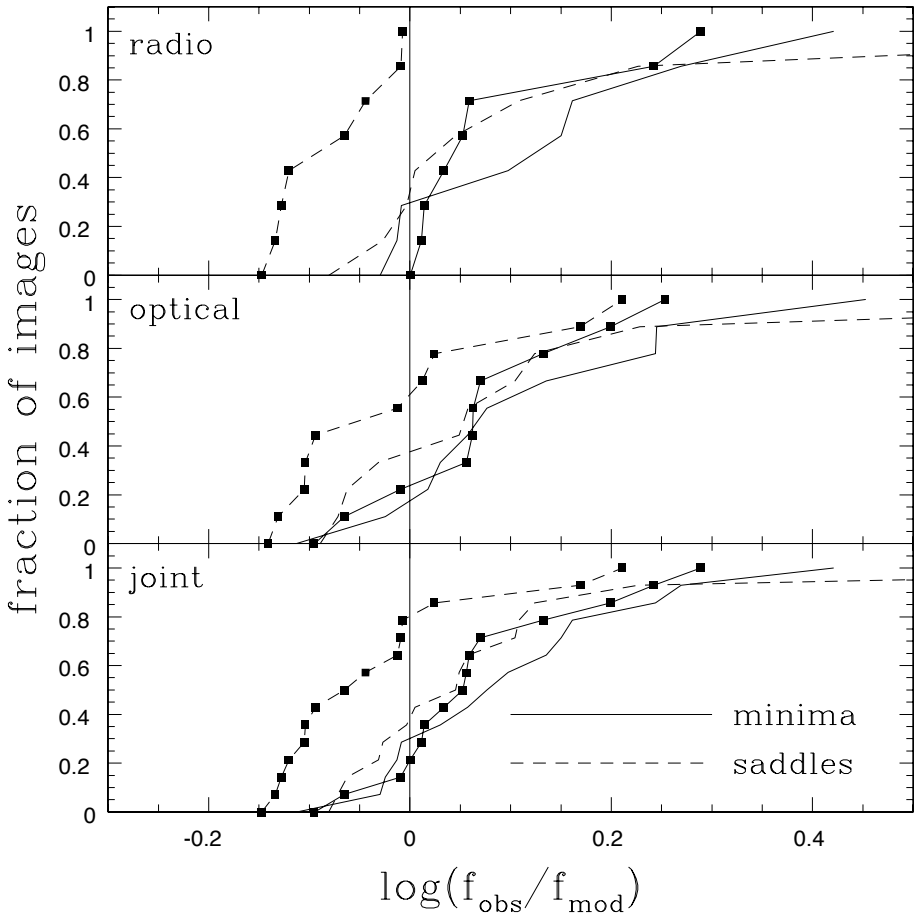


Fig. B.61. Saddle point suppression in lenses. The three panels show the cumulative distributions of model flux residuals, $\log(f_{obs}/f_{mod})$, in the real data, assuming constant fractional flux errors for each image. The solid (dashed) lines are for minima (saddle points), with squares (no squares) for the distribution corresponding to the most (least) magnified image. From top to bottom the distributions are shown for samples of 8 radio, 10 optical or 15 total four-image lenses. If the flux residuals are created by propagation effects we would not expect the distributions to depend on the image parity or magnification, while if they are due to low optical depth substructure we would expect the distribution for the brightest saddle points to be shifted to lower observed fluxes.

It turns out that anomalous flux ratios are very common – a fact which had been staring us in the face but was ignored because most people (including the author!) were mainly just annoyed that the flux ratios could not be used to constrain the potential of the primary lens so as to determine the radial mass profile. When Dalal & Kochanek (2002) collected the available four-image radio lenses to estimate the abundance of substructure, they found that 5 of 6 systems showed anomalies. In order to estimate the abundance of substructure Dalal & Kochanek (2002) developed a Bayesian Monte Carlo method which estimated the likelihood that adding substructure would significantly improve models of seven four-image lenses including the fact that the model for the primary lens would have to be adjusted each time any substructure was added. Under the assumption that the uncertainties in flux measurements (systematic as well as statistical) were 10%, they found a substructure mass fraction of $0.006 < f_{sat} < 0.07$ (90% confidence) with a median estimate of $f_{sat} = 0.02$. This is consistent with expectations from CDM simulations, including estimates of the destruction of the satellites in the inner regions of galaxies (Zentner & Bullock 2003, Mao et al. 2004), and too high to be explained by normal satellite populations. Because the result is driven by the flux anomalies, which do not depend on the mass of the substructures, rather than astrometric anomalies, which do depend on the mass, the results had almost no ability to estimate the mass scale associated with the substructure.

While substructure with approximately the surface density expected from CDM is consistent with the data, it is worth examining other possibilities. We would expect any effect from the ISM to be strongly frequency dependent (whether in the radio or in the optical). At least for radio lenses, Kochanek & Dalal (2004) found that the optical depth function needed to explain the radio flux anomalies would have to be gray, ruling out all the standard radio suspects. We would also expect propagation effects at radio frequencies to preferentially affect the faintest images because they have the smallest angular sizes – remember that more magnified images are always bigger even if you cannot resolve the change in size. The ISM also cannot discriminate between images based on parity – the ISM is a local property of the lens and the parity is not, so they cannot show a correlation. Hence, if radio propagation effects created the anomalies they should be the same for minima and saddle points and more important for the fainter than the brighter images. Fig. B.61 shows the cumulative distributions of flux residuals for radio, optical and combined four-image lens samples from Kochanek & Dalal (2004). The bright saddle point images clearly have a different distribution in each case, as we would expect for substructure but not for the ISM. The Kolmogorov-Smirnov test significance of the differences between the most magnified saddle points and the other three types of images (brightest minimum, faintest minimum, faintest saddle) is 0.04%, 5% and 0.3% for the radio, optical and joint samples respectively. The next most discrepant image is the

brightest minimum, also as expected for substructure, but with less significance. Various statistical games (bootstrap resampling methods of estimating significance or testing for anomalies) always give the same results. Thus, the ISM is ruled out as an explanation.

Even though simple Taylor series arguments make it unlikely that changes to the central potential are a solution (see §B.4.4), it still has its advocates (Evans & Witt 2003, Quadri et al. 2003, Möller, Hewitt & Blain 2003, Kawano et al. 2004). The basic answer is that it is possible to create flux anomalies by making the deviations of the central potential from ellipsoidal sufficiently large for the angular structure of the potential to change rapidly enough between nearby images to produce the necessary magnification changes. There are three basic problems with this solution (see §B.4.6 as well).

The first problem is that the required deviations from an ellipsoidal profile far too large. This is true even though the biggest survey of such models allowed image positions to shift by approximately 10 times their actual uncertainties in order to alter the image fluxes (Evans & Witt 2003) – had they forced the models to match the true astrometric uncertainties they would have needed even larger perturbations. Kochanek & Dalal (2004) found that models fitting the flux anomalies required $|a_4| \gg 0.01$ compared to the typical values observed for galaxies and simulated halos $|a_4| \sim 0.01$ (see §B.4.4). It is fair to say, however, that the quantitative results on the multipole structure of simulated halos are limited.

The second problem is that when we test these solutions in lenses for which we have additional model constraints, the models are forced back toward the standard ellipsoidal models. The basic problem, as Evans & Witt (2003) show, is that the problem of fitting image positions and fluxes with potentials of the form $rF(\theta)$ can be reduced a a problem in linear algebra if $F(\theta)$ is expanded as a multipole series – by adding enough terms it is possible to fit any four-image lens exactly. The reasons go back to the lack of constraints we discussed in §B.4.6. Fig B.27 illustrates this point using the lens B1933+503. Kochanek & Dalal (2004) first fit the four compact images with a model including deviations from an ellipsoidal surface density. With sufficiently strong deviations there were models that could eliminate the flux anomalies in this system. However, this lens, B1933+503, actually has three components to its source – a compact core forming the four-image system with the anomaly but also to radio lobes lensed into another four-image system and a two-image system for 10 images in all (Fig. B.7). When we add the constraints from these other images the model is forced back to being a standard ellipsoidal model with a flux ratio anomaly. In the future, the degree to which lens galaxy potentials are ellipsoidal could be thoroughly tested in the lenses with Einstein ring images of their host galaxies.

The third problem with using the central potential to produce flux ratio anomalies is that it does not lead to the discrimination between saddle points and minima shown in Fig. B.61. Kochanek & Dalal (2004) demonstrate this

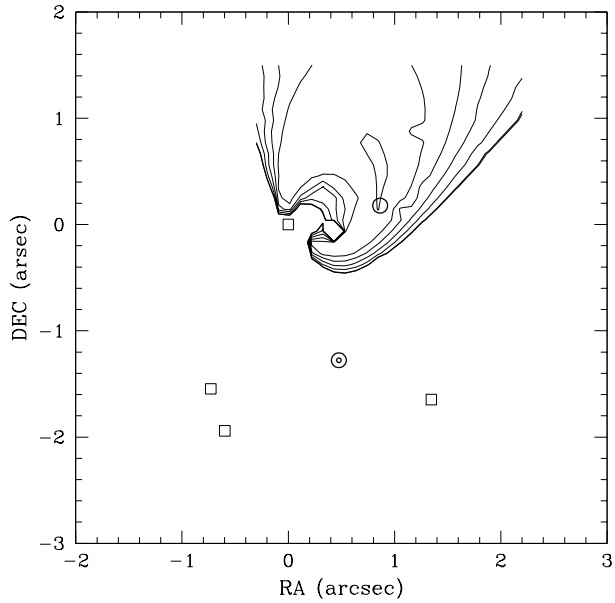


Fig. B.62. The improvement in the fit to the Ros et al. (2000) VLBI data on MG0414+0534 from adding an additional lens with a Einstein radius 15% that of the primary lens galaxy as a function of its position. The squares show the location of the quasar images, the central circles mark the position of the main lens galaxy and the single circle marks the position of object X (see Fig. B.6). The heavy contour has the same $\chi^2 = 123$ as single component models, and they then drop a factor of 0.2 per lighter contour to a minimum of $\chi^2 = 0.6$ almost exactly at the position of Object X.

with Monte Carlo simulations, but the basic reason is simple. Consider a lens like PG1115+080 with two images merging at a saddle point. The sense with which the saddle point and minima are perturbed depends on the phase of the higher order multipoles relative to the images and the critical line, but for any fixed lens potential, that phase varies depending on the source position, so the average effect cannot make the bright saddle points show a significantly different set of properties from the bright minima. Every observed flux anomaly could be explained by adding complex angular structures to the main lens, but the inability of these models to differentiate between saddle points and minima would still rule them out.

For the moment there are two barriers to improving estimates of the substructure mass fraction. First, radio lens surveys have run out of sources bright enough to conduct efficient surveys. This will only change as upgrades to existing radio arrays are completed. The proposed Merlin and VLA upgrades will provide both sensitivity and resolution improvements that will

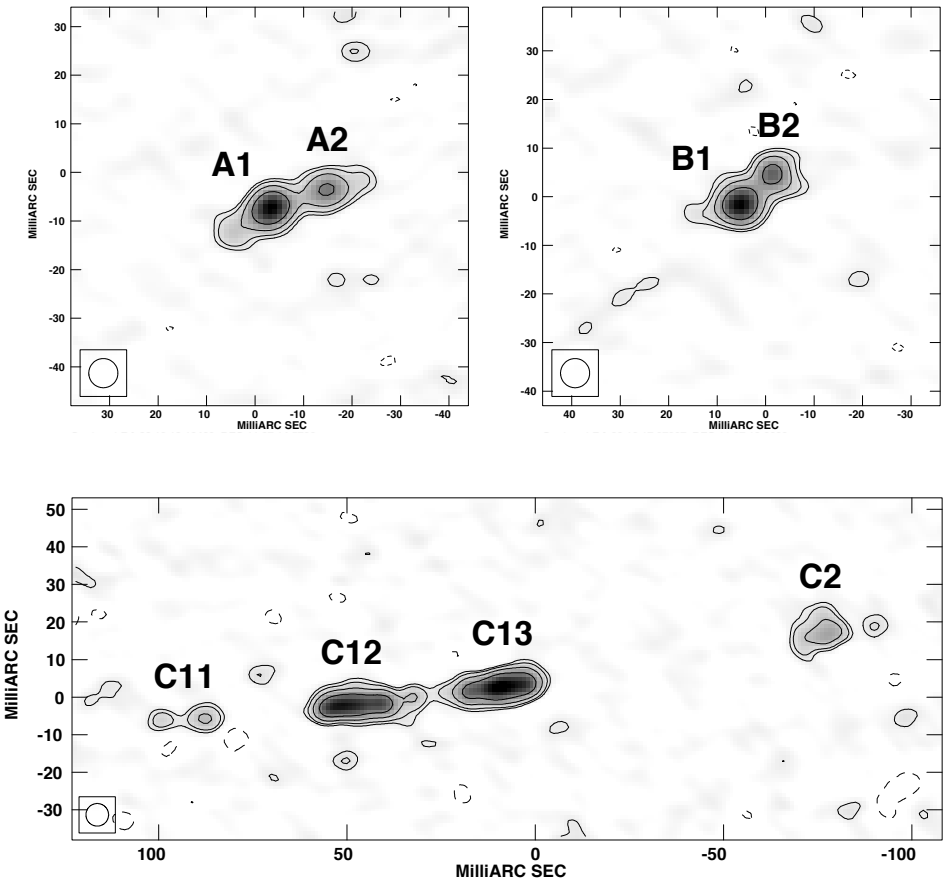


Fig. B.63. VLBI maps of MG2016+112 (Koopmans et al. 2002). The large difference in the C_{11}/C_{12} separation as compared to the C_{13}/C_2 separation is the clearest example of an "astrometric" anomaly in a lens. The critical line passes between C_{12} and C_{13} and by symmetry we would expect the separations of the subcomponents on either side of the critical line to be similar. In this case the cause of the asymmetry seems to be a galaxy D about $0\farcs8$ South of the C image (see Fig. B.6). Galaxy D has the same redshift as the primary lens (Koopmans & Treu 2002).

make the next generation of radio lens surveys easier than the last. Second, searches for substructure using optical quasars need to separate the effects of microlensing and substructure. With simple imaging this can be done by finding parts of the quasar which are sufficiently extended to avoid significant contamination from microlensing. Emission line (e.g. Moustakas & Metcalf 2003) and dust emission regions should both be large enough to filter out the effects of the stars. Studying emission line ratios is now relatively easy because of the new generation of small-pixel integral field spectrographs on 8m-class telescopes. Mid-infrared flux ratios for the dusty regions remain difficult, but the have been obtained for one lens (Q2237+0305, Agol et al. 2000) and could be extended to several more.

The gold standard, however, would be astrometric detection of dark substructure so that we would obtain a direct, mass estimate. In all the present analyses, the most massive substructures were included as part of the model. They were not, however, dark substructures because they matched to satellites visible in HST images of the lenses. For example, Object X in MG0414+0534 (Fig. B.6) has effects on the image positions that are virtually impossible to reproduce with changes in the potential of the central lens galaxy (Trotter, Winn & Hewitt 2000), while models with it easily fit the data (Ros et al. 2000). Fig. B.62 shows the dependence of the goodness of fit to MG0414+0534 on the location of an additional lens component, with a deep minimum located at the observed position of Object X. The deflections produced by an object of mass M generally scale as $M^{1/2}$, so it is relatively easy to detect the deflection perturbations from objects only 1% the mass of the primary lens. One approach is to search lenses with VLBI structures for signs of perturbations. This has been attempted for B1152+199 by Metcalf (2002), but the case for substructure is not very solid given the limited nature of the data. The cleanest example of astrometric detection of something small, but sadly not dark, is in the VLBI structure of image C in MG2016+112 (Koopmans et al. 2002). The asymmetry in the VLBI component separations of image C on either side of the critical line (see Fig. B.63) is due to a very faint galaxy $0\farcs8$ South of the image with a deflection scale $\sim 10\%$ of the primary lens (see Fig. B.6). This is in reasonable agreement with the prediction from the H-band magnitude difference of 4.6 mag and the (lens) Faber-Jackson relation between magnitudes and deflections. In this case, we even know that the satellite is at the same redshift as the lens because Koopmans & Treu (2002) accidentally measured its redshift in the course of their observations to measure the velocity dispersion of the lens galaxy.

B.8.1 Low Mass Dark Halos

When we are examining a particular lens, almost all the substructure will consist of satellites associated with the lens, with only a $\sim 10\%$ contamination from other small halos along the line-of-sight to the source (Chen, Kravtsov & Keeton 2003). However, the excess of low mass halos in CDM mass functions

relative to visible galaxies is a much more general problem because the low mass CDM satellites should exist everywhere, not just as satellites of massive galaxies. Crudely, luminosity functions diverge as $dn/dL \sim 1/L \sim 1/M$ while CDM mass functions diverge as $dn/dM \sim M^{-1.8}$ so the fraction of low mass halos that must be dark increases $\sim M^{-0.8}$ at low masses. Fig. B.51 illustrates this assuming that all low mass halos have baryons which have cooled (e.g. Gonzalez et al. 2000, Kochanek 2003c). In the context of CDM, the solution to this general problem is presumably the same as for the satellites responsible for anomalous flux ratio – they exist but lost their baryons before they could form stars. Such processes are implicit in semianalytic models which can reproduce galaxy luminosity function (e.g. Benson et al. 2003) but can be modeled empirically in much the same way was employed for the break between galaxies in clusters in §B.7 (e.g. Kochanek 2003c). In any model, the probability of the baryons cooling to form a galaxy has to drop rapidly for halo masses below $\sim 10^{11} M_\odot$ just as it has to drop rapidly for halo masses above $\sim 10^{13} M_\odot$. Unlike groups and clusters, where we still expect to be able to detect the halos from either their member galaxies or X-ray emission from the hot baryons trapped in the halo, these low mass halos almost certainly cannot be detected in emission.

We can only detect isolated, low-mass dark halos if they multiply image background sources. For SIS lenses the distribution of image separations for small separations ($\Delta\theta/\Delta\theta_* \ll 1$, Eqn. B.112) scales as

$$\frac{d\tau_{SIS}}{d\Delta\theta} \propto \Delta\theta^{1+\gamma_{FJ}(1+\alpha)/2} \tag{B.126}$$

where α describes the divergence of the mass/luminosity function at low mass and γ_{FJ} is the conversion from mass to velocity dispersion (see §B.6.2). For the standard parameters of galaxies, $\alpha \simeq -1$ and $\gamma_{FJ} \simeq 4$, the separation distribution is $d\tau_{SIS}/d\Delta\theta \propto \Delta\theta$. In practice we do not observe this distribution because the surveys have angular selection effects that prevent the detection of small image separations (below $0\farcs25$ for the radio surveys), so the observed distributions show a much sharper cutoff (Fig. B.1). Even without a cutoff, there would be few lenses to find – the CLASS survey found 9 lenses between $0\farcs3 \leq \Delta\theta \leq 1\farcs0$ in which case we expect only one lens with $\Delta\theta < 0\farcs3$ even in the absence of any angular selection effects. A VLBI survey of 3% of the CLASS sources with milli-arcsecond resolution found no lenses (Wilkinson et al. 2001), nor would it be expected to for normal galaxy populations. Our non-parametric reconstruction of the velocity function including selection effects confirms that the existing lens samples are consistent with this standard model (Fig. B.50).

The result is very different if we extrapolate to low mass with the $\alpha \simeq -1.8$ slope of the CDM halo mass function. The separation distribution becomes integrably divergent, $d\tau_{SIS}/d\Delta\theta \propto \Delta\theta^{-0.6}$, and we would expect 15 lenses with $\Delta\theta < 0\farcs3$ given 9 between $0\farcs3 \leq \Delta\theta \leq 1\farcs0$. Unfortunately, the Wilkinson et al. (2001) VLBI survey is too small to rule out such a model. A larger

VLBI survey could easily do so, allowing the lenses to confirm the galaxy counting argument for the existence of second break in the density structure of halos at low mass (Kochanek 2003c, Ma 2003) similar to the one between galaxies and high mass halos (§B.7). If the baryons in the low mass halos either fail to cool, or cool and are then ejected by feedback, then their density distributions should revert to those of their CDM halos. If they are standard NFW halos, Ma (2003) shows that such low mass dark lenses will be very difficult to detect even in far larger surveys than are presently possible. Nonetheless, improving the scale of searches for very small separations from the initial attempt by Wilkinson et al. (2001) would provide valuable limits on their existence.

The resulting small, dark lenses would be the same as the dark lenses we discussed in §B.7.2 for binary quasars and explored by Rusin (2002). They will also create the same problems about proving or disproving the lens hypothesis as was raised by the binary quasars with the added difficulty that they will be far more difficult to resolve. Time delays, while short enough to be easily measured, will also be on time scales where quasars show little variability. Confirmation of any small dark lens will probably requires systems with three or four images, rather than two images, and the presence of resolvable (VLBI) structures.

B.9 The Optical Properties of Lens Galaxies

The optical properties of lens galaxies and the properties of their interstellar medium (ISM) are important for two reasons. First, statistical calculations such as those in §B.6 rely on lens galaxies obeying the same scaling relations as nearby galaxies and the selection effects depend on the properties of the ISM. Thus, measuring the scaling relations of the observed lenses and the properties of their ISM are an important part of validating these calculations. Second, lenses have a unique advantage for studying the evolution of galaxies because they are the only sample of galaxies selected based on mass rather than luminosity, surface brightness or color. Evolution studies using optically-selected samples will always be subject to strong biases arising from the difficulty of matching nearby galaxies to distant galaxies. Selection by mass rather than light makes the lens samples almost immune to these biases.

Most lens galaxies are early-type galaxies with relatively red colors and few signs of significant on-going star formation (like the 3727Å or 5007Å Oxygen lines). The resulting need to measure absorption line redshifts is one of the reasons that the completeness of the lens redshift measurements is so poor. Locally, early-type galaxies follow a series of correlations which also exist for the lens galaxies and have been explored by Im, Griffiths & Ratnatunga (1997), Keeton, Kochanek & Falco (1998), Kochanek et al. (2000), Rusin et al. (2003), Rusin, Kochanek & Keeton (2003), van de Ven, van Dokkum & Franx (2003), Rusin & Kochanek (2004).

The first, crude correlation is the Faber-Jackson relation between velocity dispersion and luminosity used in most lens statistical calculations. A typical local relation is that from §B.6.2 and shown in Fig. B.41. Most lenses lack directly measured velocity dispersions, but all lenses have a well-determined image separation $\Delta\theta$. For specific mass models the image separation can be converted into an estimate of a velocity dispersion, such as the $\Delta\theta = 8\pi(\sigma_v/c)^2 D_{ds}/D_s$ relation of the SIS, but the precise relationship depends on the mass distribution, the orbital isotropy, the ellipticity and so forth (see §B.4.9). For the lenses, there is a close relationship between the Faber-Jackson relation and aperture mass-to-light ratios. The image separation, $\Delta\theta$, defines the aperture mass interior to the Einstein ring,

$$M_{ap} = \frac{\pi}{4}\Sigma_c \Delta\theta^2 \tag{B.127}$$

where $\Sigma_c = c^2 D_s/4\pi G D_{ds} D_d$ is the critical surface density. By image separation we usually mean either twice the mean distance of the images from the lens galaxy or twice the critical radius of a simple lens model rather than a directly measured image separation because these quantities will be less sensitive to the effects of shear and ellipticity. If we measure the luminosity in the aperture L_{ap} using (usually) HST, then we know the aperture mass-to-light (M/L) ratio $\Upsilon_{ap} = M_{ap}/L_{ap}$.

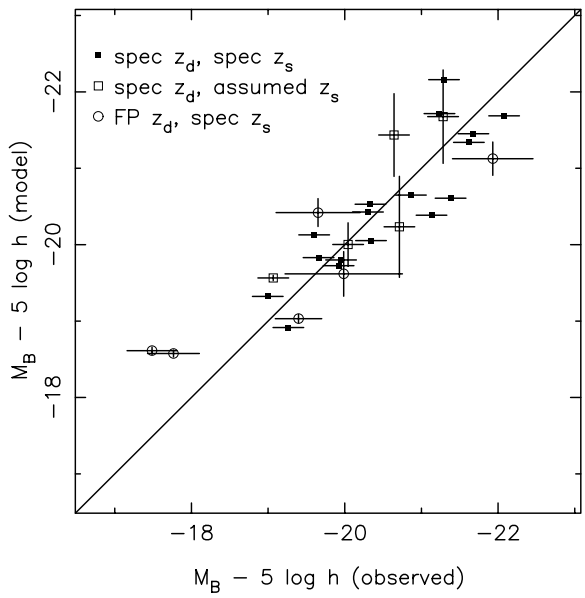
If the mass-to-light ratio varies with radius or with mass, then to compare values of Υ_{ap} from different lenses we must correct them to a common radius and common mass. If these scalings can be treated as power laws, then we can define a corrected aperture mass to light ratio $\Upsilon_* = \Upsilon_{ap}(D_d^{ang}\Delta\theta/2R_0)^x$ where R_0 is a fiducial radius and x is an unknown exponent, and we would expect to find a correlation of the form

$$\log \Upsilon_* = 2(1+a)\log\Delta\theta + 0.4 M_{abs} + \text{constant} \tag{B.128}$$

where M_{abs} is the absolute magnitude of the lens (in some band) and a value $a \neq 0$ indicates that the mass-to-light ratio varies either with mass or with radius. We can then rewrite this in a more familiar form as

$$M_{abs} = M_{abs,0} + \gamma_{EV} z_l - 1.25\gamma_{FJ}\log\left(\frac{\Delta\theta}{\Delta\theta_0}\right) \tag{B.129}$$

where $\Delta\theta_0$ sets an arbitrary separation scale, γ_{EV} (or a more complicated function) determines the evolution of the luminosity with redshift, and $\gamma_{FJ} = 4(1+a)$ sets the scaling of luminosity with normalized separation defined so that for an SIS lens (where $\Delta\theta \propto \sigma_v^2$) the exponent γ_{FJ} will match the index of the Faber-Jackson relation (Eqn. B.102). Fig. B.64 shows the resulting relation converted to the rest frame B band at redshift zero. The relation is slightly tighter than local estimates of the Faber-Jackson relation, but the scatter is still twice that expected from the measurement errors. The best fit exponent $\gamma_{FJ} = 3.29 \pm 0.58$ (Fig. B.65) is consistent with local estimates and



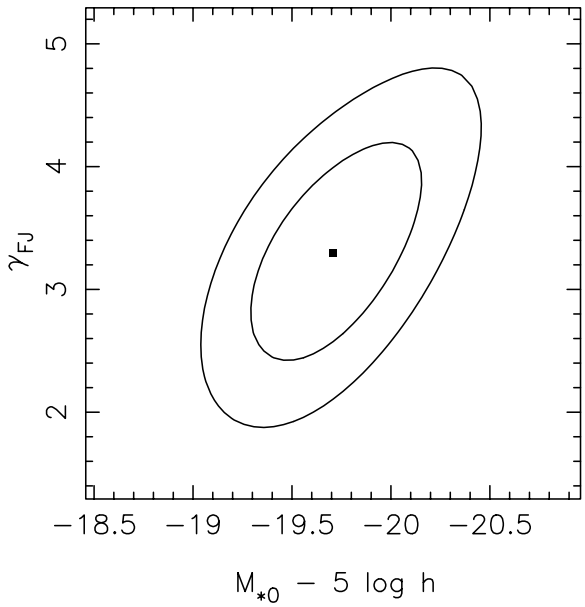


Fig. B.64. (Top) The "Faber-Jackson" relation for gravitational lenses. The figure compares the observed absolute B magnitude corrected for evolution to that predicted from the equivalent of the Faber-Jackson relation for gravitational lenses (Eqn. B.129). The different point styles indicate whether the lens and source redshifts were directly measured or estimated. From Rusin et al. (2003).

Fig. B.65. (Bottom) The redshift zero absolute B-band magnitude and effective exponent of the "Faber-Jackson" relation $L \propto \Delta\theta^{\gamma_{FJ}/2}$ for gravitational lenses.

implies a scaling exponent $a = -0.18 \pm 0.14$ that is marginally non-zero. If the mass-to-light ratio of early-type galaxies increases with mass as $\Upsilon \propto M^x$, then $x = -a = 0.18 \pm 0.14$ is consistent with estimates from the fundamental plane that more massive early-type galaxies have higher mass-to-light ratios. The solutions also require evolution with $\gamma_{EV} = -0.41 \pm 0.21$, so that early-type galaxies were brighter in the past. These scalings can also be done in terms of observed magnitudes rather than rest frame magnitudes to provide simple estimation formulas for the apparent magnitudes of lens galaxies in various bands as a function of redshift and separation to an rms accuracy of approximately 0.5 mag (see Rusin et al. 2003).

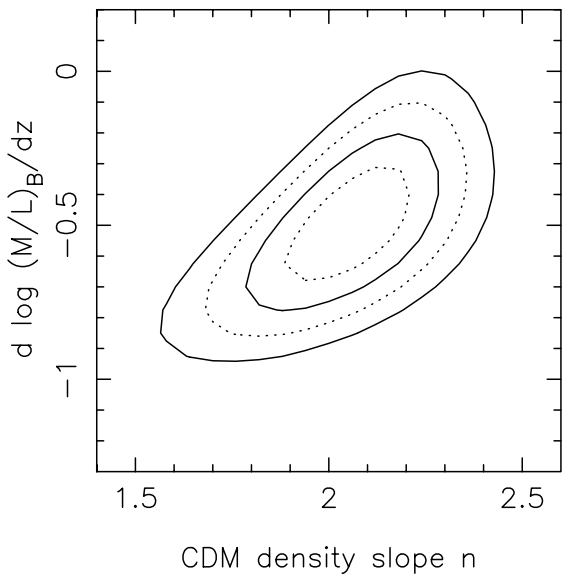
The significant scatter of the Faber-Jackson relation makes it a crude tool. Early-type galaxies also follow a far tighter correlation known as the fundamental plane (FP, Dressler et al. 1987, Djorgovski & Davis 1987) between the central, stellar velocity dispersion σ_c, the effective radius R_e and the mean surface brightness inside the effective radius $\langle \mathrm{SB}_e \rangle$ of the form

$$\log\left(\frac{R_e}{h^{-1}\mathrm{kpc}}\right) = \alpha \log\left(\frac{\sigma_c}{\mathrm{km\ s^{-1}}}\right) + \beta\left(\frac{\langle \mathrm{SB}_e \rangle}{\mathrm{mag\ arcsec^{-2}}}\right) + \gamma \qquad \text{(B.130)}$$

where the slope α and the zero-point γ depend on wavelength but the slope $\beta \simeq 0.32$ does not (e.g. Scodeggio et al. 1998, Pahre, de Carvalho & Djorgovski 1998). Local estimates for the rest frame B-band give $\alpha = 1.25$ and $\gamma_0 = -8.895 - \log(h/0.5)$ (e.g. Bender et al. 1998). In principle both the zero points and the slopes may evolve with redshift, but all existing studies have assumed fixed slopes and studied only the evolution of the zero point with redshift. For galaxies with velocity dispersion measurements, the basis of the method is that measurement of R_e and σ_v provides an estimate of the surface brightness the galaxy will have at redshift zero. The difference between the measured surface brightness at the observed redshift and the surface brightness predicted for $z = 0$ measures the evolution of the stellar populations between the two epochs as a shift in the zero-point $\Delta\gamma$. The change in the zero-point is related to the change in the luminosity by $\Delta L = -0.4\Delta\mathrm{SB}_e = \Delta\gamma/(2.5\beta)$. While these estimates are always referred to as a change in the mass-to-light ratio, no real mass measurement enters operationally. If, however, we assume a non-evolving virial mass estimate $M = c_M \sigma_v^2 R_e/G$ for some constant c_M, then the FP can be rewritten in terms of a mass-to-light ratio,

$$\log\Upsilon = \log\left(\frac{M}{L}\right) \propto \left(\frac{10\beta - 2\alpha}{5\beta}\right)\log\sigma_c + \left(\frac{2-5\beta}{5\beta}\right)\log R_e - \frac{\gamma}{2.5\beta} \qquad \text{(B.131)}$$

so that if both α and β do not evolve, the evolution of the mass-to-light ratio is $d\log\Upsilon/dz = -(d\gamma/dz)/(2.5\beta)$. Either way of thinking about the FP, either as an empirical estimator of the redshift zero surface brightness or an implicit estimate of the virial mass, leads to the same evolution estimates but alternate ways of thinking about potential systematic errors.



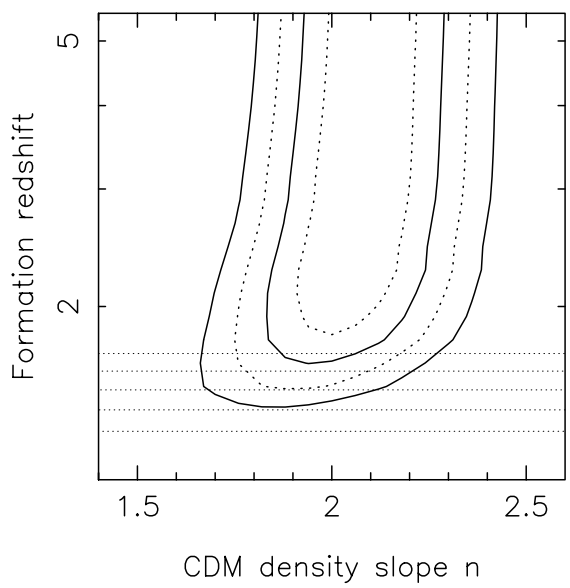


Fig. B.66. (Top) Constraints on the B-band luminosity evolution rate $d\log(M/L)_B/dz$ as a function of the logarithmic density slope n ($\rho \propto r^{-n}$) of the galaxy mass distribution. Solid (dashed) contours are the 68% and 95% confidence limits on two parameter (one parameter). These use the self-similar mass models of Eqn. B.89 and are closely related to the fundamental plane. From Rusin & Kochanek (2004).

Fig. B.67. (Bottom) Constraints on the mean star formation epoch $\langle z_f \rangle$ as a function of the logarithmic density slope n ($\rho \propto r^{-n}$) of the galaxy mass distribution. Solid (dashed) contours are the 68% and 95% confidence limits on two parameter (one parameter). The horizontal dotted lines mark $\langle z_f \rangle = 1.3$, 1.4, 1.5, 1.6 and 1.7. The lens sample favors older stellar populations with $\langle z_f \rangle > 1.5$ at 95% confidence. These use the self-similar mass models of Eqn. B.89 and are closely related to the fundamental plane. From Rusin & Kochanek (2004).

Confusion about applications of lenses to the FP and galaxy evolution usually arise because most gravitational lenses lack direct measurements of the central velocity dispersion. Before addressing this problem, it is worth considering what is done for distant galaxies with direct measurements. The central dispersion appearing in the FP has a specific definition – usually either the velocity dispersion inside the equivalent of a $3\farcs0$ aperture in the Coma cluster or the dispersion inside $R_e/8$. Measurements for particular galaxies almost never exactly match these definitions, so empirical corrections are applied to adjust the velocity measurements in the observed aperture to the standard aperture. As we explore more distant galaxies, resolution problems mean that the measurement apertures become steadily larger than the standard apertures. The corrections are made with a single, average local relation for all galaxies – implicit in this assumption is that the dynamical structure of the galaxies is homogeneous and non-evolving. This seems reasonable since the minimal scatter around the FP seems to require homogeneity, but says nothing about evolution. These are also the same assumptions used in the lensing analyses.

If early-type galaxies are homogeneous and have mass distributions that are homologous with the luminosity distributions, then there is no difference between the lens FP and the normal kinematic FP, independent of the actual mass distribution of the galaxies (Rusin & Kochanek 2004). If the mass distributions are homologous, then the mass and velocity dispersion are related by $M = c_M \sigma_c^2 R_e/G$ where c_M is a constant, σ_c is the central velocity dispersion (measured in a self-similar aperture like the $R_e/8$ aperture used in many local FP studies), and R_e is the effective radius. If we allow the mass-to-light ratio to scale with luminosity as $\Upsilon \propto L^x$, then the normal FP can be written as

$$\log R_e = \frac{2}{2x+1}\log\sigma_c + \frac{0.4(x+1)}{2x+1}\langle SB_e\rangle + \frac{\log c_M}{2x+1}, \qquad \text{(B.132)}$$

which looks like the local FP (Eqn. B.130) if $\alpha = 2/(2x+1)$ and $\beta = 0.4(x+1)/(2x+1)$ (see Faber et al. 1987). Thus, the lens galaxy FP will be indistinguishable from the FP provided early-type galaxies are homologous and the slopes can be reproduced by a scaling of the mass-to-light ratio (as they can for $x \simeq 0.3$ given $\alpha \simeq 1.2$ and $\beta \simeq 0.3$, e.g., Jorgensen, Franx & Kjaergaard 1996 or Bender et al. 1998). All the details about the mass distribution, orbital isotropies and the radius interior to which the velocity dispersion is measured enter only through the constant c_M or equivalently from differences between the FP zero point γ measured locally and with gravitational lenses. In practice, Rusin & Kochanek (2004) show that the zero point must be measured to an accuracy significantly better than $\Delta\gamma = 0.1$ before there is any sensitivity to the actual mass distribution of the lenses from the FP. Thus, there is no difference between the aperture mass estimates for the FP and its evolution and the normal stellar dynamical approach unless the major assumption underlying both approaches is violated. It also means,

perhaps surprisingly, that measuring central velocity dispersions adds almost no new information once these conditions are satisfied.

Rusin & Kochanek (2004) used the self-similar models we described in §B.4.8 to estimate the evolution rate and the star formation epoch of the lens galaxies while simultaneously estimating the mass distribution. Thus, the models for the mass include the uncertainties in the evolution and the reverse. Fig. B.66 shows the estimated evolution rate, and Fig. B.67 shows how this is related to a limit on the average star formation epoch $\langle z_f \rangle$ based on Bruzual & Charlot (1993, BC96 version) population synthesis models. This estimate is consistent with the earlier estimates by Kochanek et al. (2000) and Rusin et al. (2003) which used only isothermal lens models, as we would expect. Van de Ven, van Dokkum & Franx (2003) found a somewhat lower star formation epoch ($\langle z_f \rangle = 1.8^{1.4}_{-0.5}$) when analyzing the same data, which can be traced to differences in the analysis. First, by weighting the galaxies by their measurement errors when the scatter is dominated by systematics and by dropping two higher redshift lens galaxies with unknown source redshifts, van de Ven et al. (2003) analysis reduces the weight of the higher redshift lens galaxies, which softens the limits on low $\langle z_f \rangle$. Second, they used a power law approximation to the stellar evolution tracks which underestimates the evolution rate as you approach the star formation epoch, thereby allowing lower star formation epochs. These two effects leverage a small difference in the evolution rate[10] into a much more dramatic difference in the estimated star formation epoch. These evolution rates are consistent with estimates for cluster or field ellipticals by (e.g. van Dokkum et al. 1996, 2001, van Dokkum & Franx 2001, van Dokkum & Ellis 2003, Kelson et al. 1997, 2000), and inconsistent with the much faster evolution rates found by Treu et al. (2001, 2002) or Gebhardt et al. (2003).

B.9.1 The Interstellar Medium of Lens Galaxies

As well as studying the emission by the lens galaxy we can study its absorption of emission from the quasar as a probe of the interstellar medium (ISM) of the lens galaxies. The most extensively studied effect of the ISM is dust extinction because of its effects on estimating the cosmological model from optically-selected lenses and because it allows unique measurements of extinction curves outside the local Group. There are also broad band effects on the radio continuum due to free-free absorption, scatter broadening and Faraday rotation. While all three effects have been observed, they have been of little practical importance so far. Finally, in both the radio and the optical, the lens can introduce narrow absorption features. While these are observed

[10] Rusin & Kochanek (2004) obtained $d\log(M/L)_B/dz = -0.50 \pm 0.19$ including the uncertainties in the mass distribution, Rusin et al. (2003) obtained -0.54 ± 0.09 for a fixed SIS model, and van de Ven et al. (2003) obtained -0.62 ± 0.13 for a fixed SIS model.

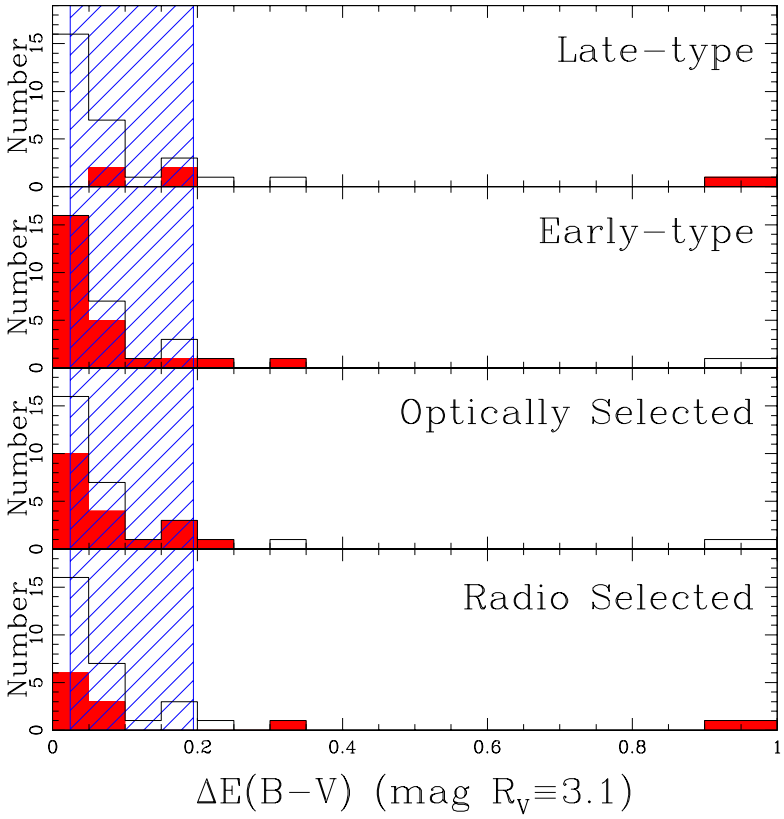


Fig. B.68. Histograms of the differential extinction in various lens subsamples from Falco et al. (1999). In each panel the solid histogram shows the full sample of 37 differential extinctions measured in 23 lenses while the shaded histogram shows the distributions for different selection methods (radio/optical) or galaxy types (early/late). The hatched region shows the extinction range consistent with the Falco, Kochanek & Muñoz (1998) analysis of the difference between the statistics of radio-selected and optically-selected lens samples (see §B.6.6). Note that the most highly extincted systems, PKS1830–211 and B0218+357, are both radio-selected and late-type galaxies. The lowest differential extinction bins are contaminated by the effects of finite measurement errors.

in some lenses, observational limitations have prevented them from being as useful as the are in other areas of astrophysics.

As we mentioned in §B.6, extinction is an important systematic problem for estimating the cosmological model using the statistics of optically selected lenses. It modifies the results by changing the effective magnification bias of the sample because it can make lensed quasars dimmer than their unlensed counterparts. Because we see multiple images of the same quasar, it is relatively easy to estimate the differential extinction between lensed images under the assumption that the quasar spectral shapes are not varying on the time scale corresponding to the time delay between the images and that microlensing effects are not significantly changing the slope of the quasar continuum. The former is almost certainly valid, while for the latter we simply lack the necessary data to check the assumption (although we have a warning sign from the systems where the continuum and emission line flux ratios differ, see Part 4). Under these assumptions, the magnitude difference at wavelength λ between two images A and B

$$m_A(\lambda) - m_B(\lambda) = -2.5 \log \left| \frac{\mu_A}{\mu_B} \right| + R\left(\frac{\lambda}{1+z_l}\right) \Delta E(B-V) \qquad \text{(B.133)}$$

depends on the ratio of the image magnifications μ_A/μ_B, the differential extinction $\Delta E(B-V) = E_A - E_B$ between the two images and the extinction law $R(\lambda/(1+z_l))$ of the dust in the rest frame of the dust. We have the additional assumption that either the extinction law is the same for both images or that one image dominates the total extinction (Nadeau et al. 1991). Because it is a purely differential measurement that does not depend on knowing the intrinsic spectrum of the quasar, it provides a means of determining extinctions and extinction laws that is otherwise only achievable locally where we can obtain spectra of individual stars (the pair method, e.g. Cardelli, Clayton & Mathis 1989). The total extinction cannot be determined to any comparable accuracy because estimates of the total extinction require an estimate of the intrinsic spectrum of the quasar. Fig. B.68 shows the distribution of differential extinctions found in the Falco et al. (1999) survey of extinction in 23 gravitational lenses. Only 7 of the 23 systems had colors consistent with no extinction, and after correcting for measurement errors and excluding the two outlying, heavily extincted systems the data are consistent with a one-sided Gaussian distribution of extinctions starting at 0 and with a dispersion of $\sigma_{\Delta E} \simeq 0.1$ mag. The two outlying systems, B0218+357 and PKS1830–211, were both radio-selected and both have one image that lies behind a molecular cloud of a late type lens galaxy (see below).

For lenses that have the right amount of dust, so that the image flux ratio can be measured accurately over a broad range of wavelengths, it is possible to estimate the extinction curve $R(\lambda/(1+z_l))$ of the dust (Nadeau et al. 1991) or to estimate the dust redshift under the assumption that the extinction curve is similar to those measured locally (Jean & Surdej 1998).

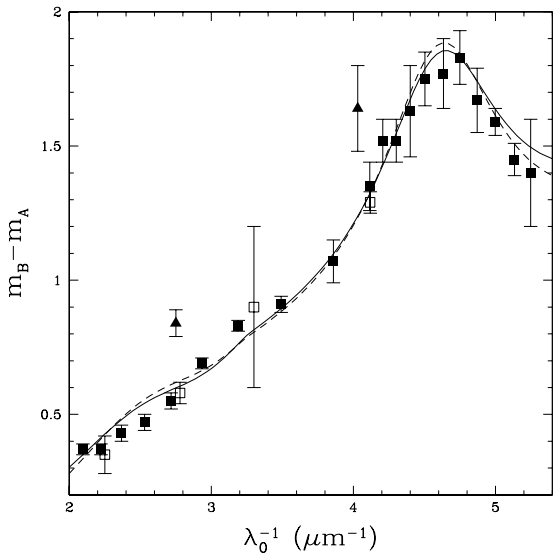


Fig. B.69. The extinction curve of the dust in SBS0909+532 at $z_l = 0.83$ by Motta et al. (2002). The solid squares show the magnitude difference as a function of inverse rest wavelength derived from integral field spectra of the continuum of the quasars. The open squares are broad band measurements from earlier HST imaging and the filled triangles are the flux ratios in the quasar emission lines. The solid curve shows the best fit $R_V = 2.1 \pm 0.9$ Cardelli, Clayton & Mathis (1989) extinction curve while the dashed curve shows a standard $R_V = 3.1$ curve. The offset between the continuum and emission line flux ratios seems not to depend on wavelength and is probably due to microlensing.

Starting with Nadeau et al. (1991), there have been many estimates of extinction curves in lens galaxies (Falco et al. 1999, Toft, Hjorth & Burud 2000, Motta et al. 2002, Muñoz et al. 2004). The most interesting of these are for systems where the region near the 2175Å extinction feature is visible. This requires source and lens redshifts that put the feature at long enough wavelengths to be easily observed (i.e. higher lens redshifts) with a quasar UV continuum extending to shorter wavelengths (i.e. lower source redshifts). Motta et al. (2002) achieved the first cosmological detection of the feature in the $z_l = 0.83$ lens SBS0909+532, as shown in Fig. B.69. The overall extinction curve is marginally consistent with a standard Galactic $R_V = 3.1$ extinction curve. Other cosmologically distant extinction curves are very different from normal Galactic models ranging for an anomalously low R_V curve in MG0414+0534 at $z_l = 0.96$ (Falco et al. 1999), probably an SMC extinction curve in LBQS1009–252 at an estimated redshift of $z_l \simeq 0.88$ (Muñoz

et al. 2004), and a anomalously high R_V extinction curve for the dust in the molecular cloud of the $z_l = 0.68$ lens galaxy in B0218+357. The Jean & Surdej (1998) idea of using the shape of the extinction curve to estimate the redshift of the dust also seems to work given a reasonable amount of dust and wavelength coverage (see Falco et al. 1999, Muñoz et al. 2004), but too few lenses with unknown redshifts satisfy the requirements for widespread use of the method.

For broad band radio emission from the source, the three observed propagation effects are free-free absorption, scatter broadening and Faraday rotation. For example, in PMNJ1632–0033, the candidate third image of the lens (C) has the same radio spectrum as the other two images except at the lowest frequency observed (1.4 GHz) where it is fainter than expected. This can be interpreted as free-free absorption by electrons at the center of the lens galaxy but the interpretation needs to be confirmed by measurements at additional frequencies to demonstrate that the dependence of the optical depth on wavelength is consistent with the free-free process (Winn, Rusin & Kochanek 2004). Scatter broadening is observed in many radio lenses (e.g. PMN0134–0931, Winn et al. 2003; B0128+437, Biggs et al. 2004; PKS1830–211, Jones et al. 1996; B1933+503, Marlow et al. 1999) primarily as changes in the fluxes of images between high resolution VLBI observations and lower resolution VLA observations or apparently finite sizes for compact source components in VLBI observations. In the presence of a magnetic field, the scattering medium will also rotate polarization vectors (e.g. MG1131+0456, Chen & Hewitt 1993). This is only of practical importance if maps which depend on the polarization vector are used to constrain the lens potential. In short, these effects are observed but have so far been of little practical consequence.

More surprisingly, absorption by atoms and molecules has also been of little practical import for lens physics as yet. Wiklind & Alloin (2002) provide an extensive review of molecular absorption and emission in gravitational lenses. The two systems with the strongest absorption systems are B0218+357 and PKS1830–211 (see Gerin et al. 1997 and references therein) where one of the two images lies behind a molecular cloud of the spiral galaxy lens. These two systems also show the highest extinction of any lensed images (Falco et al. 1999). Molecular absorption systems can be used to determine time delays (Wiklind & Alloin 2002), measure the redshift of lens galaxies (the lens redshift in PKS1830-211 is measured using molecular absorption lines, Wiklind & Combes 1996), and potentially to determine the rotation velocity of the lens galaxy (e.g. Koopmans & de Bruyn 2003). These studies at centimeter and millimeter wavelengths are heavily limited by the resolution and sensitivity of existing instruments, and the importance of these radio absorption features will probably rise dramatically with the completion of the next generation of telescopes (e.g. ALMA, LOFAR, SKA).

Similar problems face studies of metal absorption lines in the optical. Since most lenses are at modest redshifts, the strongest absorption lines expected from the lens galaxies tend to be observable only from space because they lie at shorter wavelengths than the atmospheric cutoff. For most lenses only the MgII (2800Å) lines are observable from the ground since you only require a lens redshift $z_l \gtrsim 0.26$ to get the redshifted absorption lines longwards of 3500Å. The other standard metal line, CIV (1549Å), is only visible for $z_l \gtrsim 1.25$, and we have no confirmed lens redshifts in this range. Spectroscopy with HST can search for metal lines in the UV, but the integration times tend to be prohibitively long unless the quasar images are very bright. Thus, while absorption lines either associated with the lens galaxy or likely to be associated with the lens galaxy are occasionally found (e.g. SDSS1650+4251, Morgan, Snyder and Reens 2003; or HE1104–1805, Lidman et al. 2000), there have been no systematic studies of metal absorption in gravitational lenses. Nonetheless, some very bright quasar lenses are favored targets for very high dispersion studies of they Lyα forest, particularly the four-image lens B1422+231 and the three image lens APM08279+5255, because the lens magnification makes these systems anomalously bright for quasars at $z_s > 3$.

B.10 Extended Sources and Quasar Host Galaxies

As we saw in Figs. B.3, B.4, and B.8, we frequently see lensed emission from extended components of the source. These arcs and rings are important because they can supply the extra constraints needed to determine the radial mass distribution that we lack in a simple two-image of four-image lens (§B.4.1). The magnification produced by gravitational lensing also allows us to study far fainter quasar host galaxies than is otherwise possible. Comparisons of the luminosities and colors of high and low redshift host galaxies and the relative luminosities of the host and the quasar are important for understanding the growth of supermassive black holes and their relationships with their parent halos.

Modeling extended emission is more difficult than modeling point sources largely because of the complications introduced by the finite resolution of the observations. In this section we first discuss a simple theory of Einstein ring images, then some methods for modeling extended emission, and finally some results about the mass distributions of lenses and the properties of quasar host galaxies. All models of extended lenses sources start from the fact that lensing preserves the surface brightness of the source – what we perceive as magnification is only an artifact of the finite resolution of our observations. This can be modified by absorption in the ISM of the lens galaxy (e.g. see, Koopmans et al. 2003), but we will neglect this complication in what follows. We start with a simple analytic model for the formation of Einstein rings, then discuss numerical reconstructions of lensed sources and their ability

to constrain mass distributions, and end with a survey of the properties of quasar host galaxies.

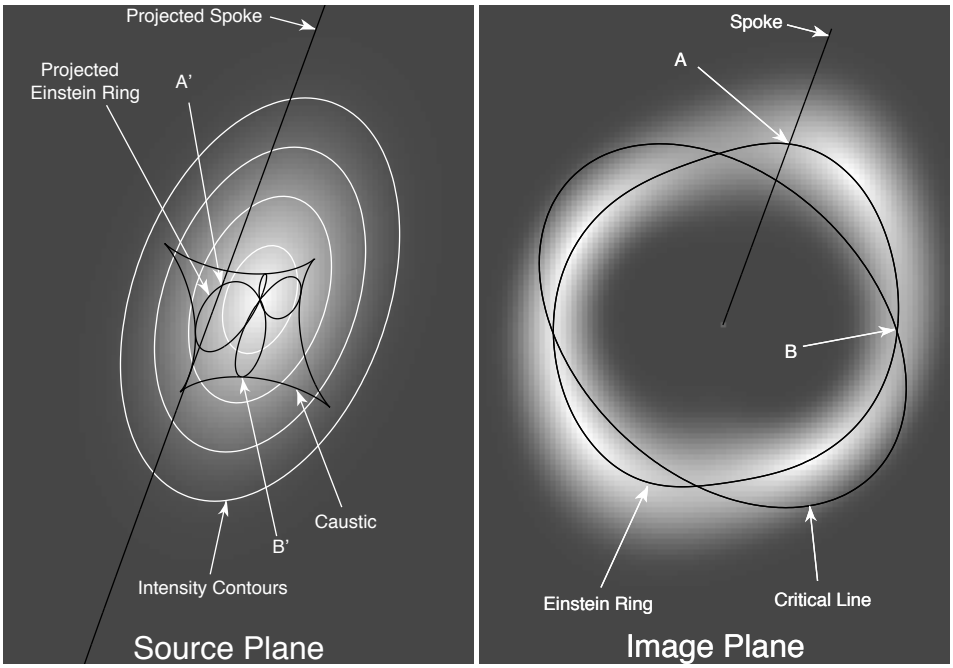


Fig. B.70. An illustration of ring formation by an SIE lens. An ellipsoidal source (left gray-scale) is lensed into an Einstein ring (right gray-scale). The source plane is magnified by a factor of 2.5 relative to the image plane. The tangential caustic (astroid on left) and critical line (right) are superposed. The Einstein ring curve is found by looking for the peak brightness along radial spokes in the image plane. For example, the spoke in the illustration defines point A on the ring curve. The long line segment on the right is the projection of the spoke onto the source plane. Point A corresponds to point A' on the source plane where the projected spoke is tangential to the intensity contours of the source. The ring in the image plane projects into the four-lobed pattern on the source plane. Intensity maxima along the ring correspond to the center of the source. Intensity minima along the ring occur where the ring crosses the critical curve (e.g. point B). The corresponding points on the source plane (e.g. B') are where the astroid caustic is tangential to the intensity contours.

B.10.1 An Analytic Model for Einstein Rings

Most of the lensed extended sources we see are dominated by an Einstein ring – this occurs when the size of the source is comparable to the size of

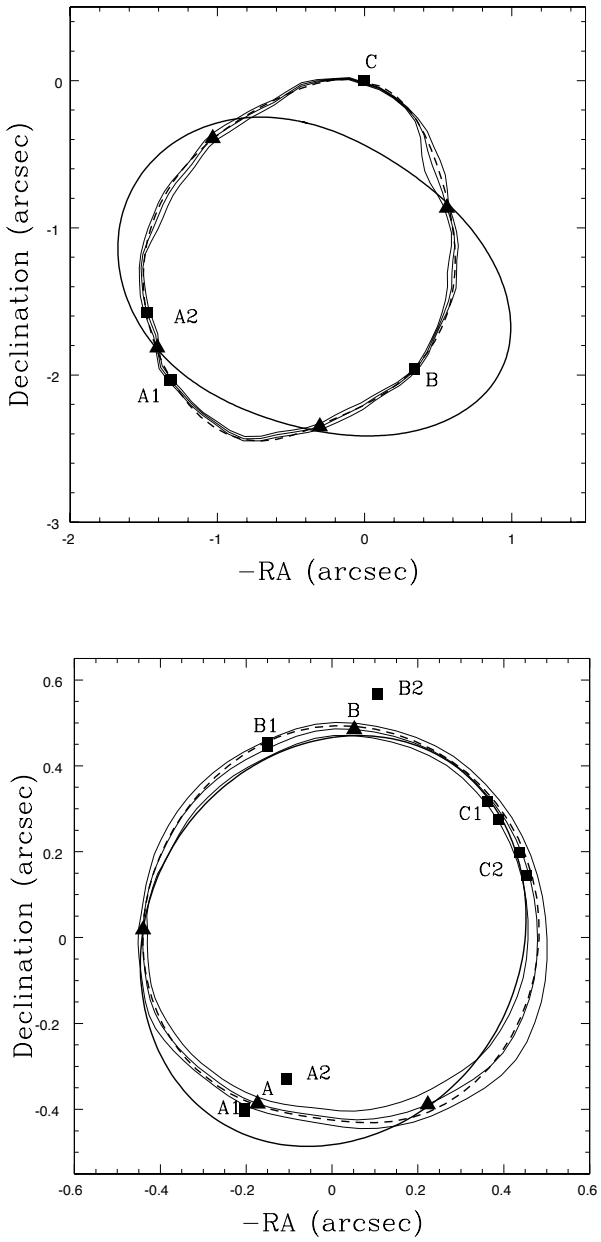


Fig. B.71. The Einstein ring curves in PG1115+080 (top) and B1938+666 (bottom). The black squares mark the lensed quasar or compact radio sources. The light black lines show the ring curve and its uncertainties. The black triangles show the intensity minima along the ring curve (but not their uncertainties). The best fit model ring curve is shown by the dashed curve, and the heavy solid curve shows the critical line of the best fit model. The model was not constrained to fit the critical line crossings.

the astroid caustic associated with producing four-image lenses. When the Einstein ring is fairly thin, there is a simple analytic model for the formation of Einstein rings (Kochanek, Keeton & McLeod 2001). The important point to understand is that the ring is a pattern rather than a simple combination of multiple images. Mathematically, what we identify as the ring is the peak of the surface brightness as a function of angle around the lens galaxy. We can identify the peak by finding the maximum intensity $\lambda(\chi)$ along radial spokes in the image plane, $\boldsymbol{\theta}(\lambda) = \boldsymbol{\theta}_0 + \lambda(\cos\chi, \sin\chi)$. At a given azimuth χ we find the extremum of the surface brightness of the image $f_D(\boldsymbol{\theta})$ along each spoke, and these lie at the solutions of

$$0 = \partial_\lambda f_D(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} f_D(\boldsymbol{\theta}) \cdot \frac{d\boldsymbol{\theta}}{d\lambda}. \tag{B.134}$$

The next step is to translate the criterion for the ring location onto the source plane. In real images, the observed image $f_D(\boldsymbol{\theta})$ is related to the actual surface density $f_I(\boldsymbol{\theta})$ by a convolution with the beam (PSF), $f_D(\boldsymbol{\theta}) = B * f_I(\boldsymbol{\theta})$, but for the moment we will assume we are dealing with a true surface brightness map. Under this assumption $f_D(\boldsymbol{\theta}) = f_I(\boldsymbol{\theta}) = f_S(\boldsymbol{\beta})$ because of surface brightness conservation. When we change variables, the criterion for the peak brightness becomes

$$0 = \nabla_{\boldsymbol{\beta}} f_S(\boldsymbol{\beta}) \cdot M^{-1} \cdot \frac{d\boldsymbol{\theta}}{d\lambda} \tag{B.135}$$

where the inverse magnification tensor $M^{-1} = d\boldsymbol{\theta}/d\boldsymbol{\beta}$ is introduced by the variable transformation. Geometrically we must find the point where the tangent vector of the curve, $M^{-1} \cdot d\boldsymbol{\theta}/d\lambda$ is perpendicular to the local gradient of the surface brightness $\nabla_{\boldsymbol{\beta}} f_S(\boldsymbol{\beta})$. These steps are illustrated in Fig. B.70.

This result is true in general but not very useful. We next assume that the source has ellipsoidal surface brightness contours, $f_S(m^2)$, with $m^2 = \Delta\boldsymbol{\beta} \cdot S \cdot \Delta\boldsymbol{\beta}$ where $\Delta\boldsymbol{\beta} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$ is the distance from the center of the source, $\boldsymbol{\beta}_0$, and the matrix S is defined by the axis ratio $q_s = 1 - \epsilon_s \leq 1$ and position angle χ_s of the source. We must assume that the surface brightness declines monotonically, $df_s(m^2)/dm^2 < 0$, but require no additional assumptions about the actual profile. With these assumptions the Einstein ring curve is simply the solution of

$$0 = \Delta\boldsymbol{\beta} \cdot S \cdot \mu^{-1} \cdot \frac{d\boldsymbol{\theta}}{d\lambda}. \tag{B.136}$$

The ring curve traces out a four (two) lobed cloverleaf pattern when projected on the source plane if there are four (two) images of the center of the source (see Fig. B.70). These lobes touch the tangential caustic at their maximum ellipsoidal distance from the source center, and these cyclic variations in the ellipsoidal radius produce the brightness variations seen around the ring. The surface brightness along the ring is defined by $f_I(\lambda(\chi), \chi)$ for a spoke at azimuth χ and distance $\lambda(\chi)$ found by solving Eqn. B.135. The extrema

in the surface brightness around the ring are located at the points where $\partial_\chi f_I(\lambda(\chi),\chi) = 0$, which occurs only at extrema of the surface brightness of the source (the center of the source, $\Delta\boldsymbol{\beta} = 0$ in the ellipsoidal model), or when the ring crosses a critical line of the lens and the magnification tensor is singular ($|M|^{-1} = \mu^{-1} = 0$) for the minima. These are general results that do not depend on the assumption of ellipsoidal symmetry.

For an SIE lens in an external shear field we can derive some simple properties of Einstein rings to lowest order in the various axis ratios. Let the SIE have critical radius b, axis ratio $q_l = 1 - \epsilon_l$ and put its major axis along θ_1. Let the external shear have amplitude γ and orientation θ_γ. We let the source be an ellipsoid with axis ratio $q_s = 1 - \epsilon_s$ and a major axis angle χ_s located at position $(\beta\cos\chi_0, \beta\sin\chi_0)$ from the lens center. The tangential critical line of the lens lies at radius

$$r_{crit}/b = 1 + \frac{\epsilon_l}{2}\cos 2\chi - \gamma\cos 2(\chi - \chi_\gamma) \tag{B.137}$$

while the Einstein ring lies at

$$\frac{r_E}{b} = 1 + \frac{\beta}{b}\cos(\chi - \chi_0) - \frac{\epsilon_l}{6}\cos 2\chi + \gamma\cos 2(\chi - \chi_\gamma). \tag{B.138}$$

At this order, the Einstein ring is centered on the source position rather than the lens position. The orientation of the ring is generally perpendicular to that of the critical curve, although it need not be exactly so when the SIE and the shear are misaligned due to the differing coefficients of the shear and ellipticity terms in the two expressions. These results lead to a false impression that the results do not depend on the shape of the source. In making the expansion we assumed that all the terms were of the same order ($\beta/b \sim \gamma \sim \epsilon_l \sim \epsilon_s$), but we are really doing an expansion in the ellipticity of the potential of the lens $e_\Psi \sim e_l/3$ rather than the ellipticity of the density distribution of the lens, so second order terms in the shape of the source are as important as first order terms in the ellipticity of the potential. For example in a circular lens with no shear ($\epsilon_l = 0$, $\gamma = 0$) the ring is located at

$$\frac{r_E}{b} = 1 + \frac{\beta}{b}\frac{(2-\epsilon_s)\cos(\chi - \chi_0) + \epsilon_s\cos(2\chi_s - \chi - \chi_0)}{2 - \epsilon_s + \epsilon_s\cos 2(\chi_s - \chi)} \tag{B.139}$$

which has only odd terms in its multipole expansion and converges slowly for flattened sources. In general, the ring shape is a weak function of the source shape only if the potential is nearly round and the source is almost centered on the lens. The structure of the lens potential dominates the even multipoles of the ring shape, while the structure of the source dominates the odd multipoles.

In fact, the shape of the ring can be used to simply "read off" the amplitudes of the higher order multipoles of the lens potential. This is nicely illustrated by an isothermal potential with arbitrary angular structure, $\Psi =$

$rbF(\chi)$ with $\langle F(\chi)\rangle = 1$ (see Zhao & Pronk 2001, Witt et al. 2000, Kochanek et al. 2001, Evans & Witt 2001) in the absence of any shear. The tangential critical line of the lens is

$$\frac{r_{crit}}{b} = F(\chi) + F''(\chi). \tag{B.140}$$

If $\hat{e}_\chi$ and $\hat{e}_\theta$ are radial and tangential unit vectors relative to the lens center and $\boldsymbol{\beta}_0$ is the distance of the source from the lens center, then the Einstein ring curve is

$$\frac{r_E}{b} = F(\chi) + F'(\chi)\frac{\hat{e}_\chi \cdot S \cdot \hat{e}_\theta}{\hat{e}_\theta \cdot S \cdot \hat{e}_\theta} + \frac{\boldsymbol{\beta}_0 \cdot S \cdot \hat{e}_\theta}{\hat{e}_\theta \cdot S \cdot \hat{e}_\theta} \to F(\chi) + \boldsymbol{\beta}_0 \cdot \hat{e}_\theta \tag{B.141}$$

with the limit showing the result for a circular source.

Thus, by analyzing the multipole structure of the ring curve one can deduce the multipole structure of the potential. While this has not been done non-parametrically, the ability of standard ellipsoidal models to reproduce ring curves strongly suggests that higher order multipoles cannot be significantly different from the ellipsoidal scalings. Fig. B.71 shows two examples of fits to the ring curves in PG1115+080 and B1938+666 using SIE plus external shear lens models. The major systematic problem with fitting the real data are that bright quasar images must frequently be subtracted from the image before the ring curve can be extracted, and this can lead to artifacts like the wiggle in the curve between the bright A_1/A_2 images of PG1115+080. Other than that, the accuracy with which the ellipsoidal (plus shear) models reproduce the curves is consistent with the uncertainties. In both cases the host galaxy is relatively flat ($q_s = 0.58 \pm 0.02$ for PG1115+080 and 0.62 ± 0.14 for B1938+666). The flatness of the host explains the "boxiness" of the PG1115+080 ring, while the B1938+666 host galaxy shape is poorly constrained because the center of the host is very close to the center of the lens galaxy so the shape of the ring is insensitive to the shape of the source. Unless the source is significantly offset from the center of the lens, as we might see for the host galaxy of an asymmetric two-image lens, it does not constrain the radial density profile of the lens very well – after considerable algebraic effort you can show that the dependence on the radial structure scales as $|\Delta\boldsymbol{\beta}|^4$. It can, however, help considerably in this circumstance because it eliminates the angular degrees of freedom in the potential that make it impossible for two-image lenses to constrain the radial density profile at all.

B.10.2 Numerical Models of Extended Lensed Sources

Obviously the ring curve and its extrema are an abstraction of the real structure of the lensed source. Complete modeling of extended sources requires a real model for the surface brightness of the source. In many cases it is sufficient to simply use a parameterized model for the source, but in other cases

it is not. The basic idea in any non-parametric method is that there is an optimal estimate of the source structure for any given lens model. This is most easily seen if we ignore the smearing of the image by the beam (PSF) and assume that our image is a surface brightness map. Since surface brightness is conserved by lensing, $f_I(\boldsymbol{\theta}) = f_S(\boldsymbol{\beta})$. For any lens model with parameters $\boldsymbol{p}$, the lens equations define the source position $\boldsymbol{\beta}(\boldsymbol{\theta}, \boldsymbol{p})$ associated with each image position. If we had only single images of each source point, this would be useless for modeling lenses. However, in a multiply imaged region, more than one point on the image plane is mapped to the same point on the source plane. In a correct lens model, all image plane points mapped to the same source plane position should have the same surface brightness, while in an incorrect model, points with differing surface brightnesses will be mapped to the same source point. This provided the basis for the first non-parametric method, sometimes known as the "Ring Cycle" method (Kochanek et al. 1989, Wallington, Kochanek & Koo 1995). Suppose source plane pixel j is associated with image plane pixels $i = 1 \cdots n_j$ with surface brightness f_i and uncertainties σ_i. The goodness of fit for this source pixel is

$$\chi_j^2 = \sum_{i=1}^{n_j} \left(\frac{f_i - f_s}{\sigma_i} \right)^2 \tag{B.142}$$

where f_s will be our estimate of the surface brightness on the source plane. For each lens model we compute $\chi^2(\boldsymbol{p}) = \sum \chi_j^2$ and then optimize the lens parameters to minimize the surface brightness mismatches.

The problem with this algorithm is that we never have images that are true surface brightness maps – they are always the surface brightness map convolved with some beam (PSF). We can generalize the simple algorithm into a set of linear equations. Although the source and lens plane are two-dimensional, the description is simplified if we simply treat them as a vector $\boldsymbol{f}_S$ of source plane surface brightness and a vector $\boldsymbol{f}_I$ of image plane flux densities (i.e. including any convolution with the beam). The two images are related by a linear operator $A(\boldsymbol{p})$ that depends on the parameters of the current lens model and the PSF. In the absence of a lens, A is simply the real-space (PSF) convolution operator. In either case, the fit statistic

$$\chi^2 = \frac{|\boldsymbol{f}_I - A(\boldsymbol{p})\boldsymbol{f}_S|^2}{\sigma^2} \tag{B.143}$$

(with uniform uncertainties here, but this is easily generalized) must first be solved to determine the optimal source structure for a given lens model and then minimized as a function of the lens model. The optimal source structure $d\chi^2/d\boldsymbol{f}_S = 0$ leads to the equation that $\boldsymbol{f}_S = A^{-1}(\boldsymbol{p})\boldsymbol{f}_I$. The problem, which is the same as we discussed for non-parametric mass models in §B.4.7, is that a sufficiently general source model when combined with a PSF will lead to a singular matrix for which $A(\boldsymbol{p})^{-1}$ is ill-defined – physically, there will be wildly oscillating source models for which it is possible to obtain $\chi^2(\boldsymbol{p}) = 0$.

Three approaches have been used to solve the problem. The first is LensClean (Kochanek & Narayan 1992, Ellithorpe, Kochanek & Hewitt 1996, Wucknitz 2004), which is based on the Clean algorithm of radio astronomy. Like the normal Clean algorithm, LensClean is a non-linear method using a prior that radio sources can be decomposed into point sources for determining the structure of the source. The second is LensMEM (Wallington, Kochanek & Narayan 1996), which is based on the Maximum Entropy Method (MEM) for image processing. The determination of the source structure is stabilized by minimizing $\chi^2 + \lambda \int d^2\beta f_S \ln(f_S/f_0)$ while adjusting the Lagrange multiplier λ such that at the minimum $\chi^2 \sim N_{dof}$ where N_{dof} is the number of degrees of freedom in the model. Like Clean/LensClean, MEM/LensMEM is a non-linear algorithm in which solutions must be solved iteratively. Both LensClean and LensMEM can be designed to produce only positive-definite sources. The third approach is linear regularization where the source structure is stabilized by minimizing $\chi^2 + \lambda \boldsymbol{f}_S \cdot H \cdot \boldsymbol{f}_S$ (Warren & Dye 2003, Koopmans et al. 2003). The simplest choice for the matrix H is the identity matrix, in which case the added criterion is to minimize the sum of the squares of the source flux. More complicated choices for H will minimize the gradients or curvature of the source flux. The advantage of this scheme is that the solution is simply a linear algebra problem with $(A^T(\boldsymbol{p})A(\boldsymbol{p}) + \lambda H)\boldsymbol{f}_S = A^T(\boldsymbol{p})\boldsymbol{f}_I$.

In all three of these methods there are two basic systematic issues which need to be addressed. First, all the methods have some sort of adjustable parameter – the Lagrange multiplier λ in LensMEM or the linear regularization methods and the stopping criterion in the LensClean method. As the lens model changes, the estimates of the parameter errors will be biased if the treatment of the multiplier or the stopping criterion varies with changes in the lens model in some poorly understood manner. Second, it is difficult to work out the accounting for the number of degrees of freedom associated with the model for the source when determining the significance of differences between lens models. Both of these problems are particularly severe when comparing models where the size of the multiply imaged region depends on the lens model. Since only multiply imaged regions supply any constraints on the model, one way to improve the goodness of fit is simply to shrink the multiply imaged region so that there are fewer constraints. Since changes in the radial mass distribution have the biggest effect on the multiply imaged region, this makes estimates of the radial mass distribution particularly sensitive to controlling these biases. It is fair to say that all these algorithms lack a completely satisfactory understanding of this problem. For radio data there are added complications arising from the nature of interferometric observations, which mean that good statistical models must work with the raw visibility data rather than the final images (see Ellithorpe et al. 1996).

These methods, including the effects of the PSF, have been applied to determining the mass distributions in 0047–2808 (Dye & Warren 2003), B0218+357 (Wucknitz, Biggs & Browne 2004), MG1131+0456 (Chen, Kochanek & He-

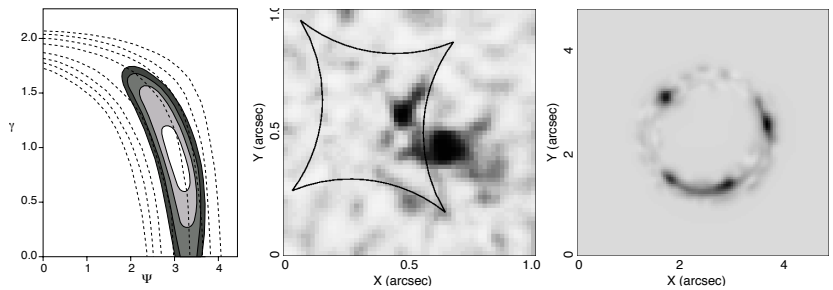


Fig. B.72. Models of 0047–2808 from Dye & Warren (2003). The right panel shows the lensed image of the quasar host galaxy after the foreground lens has been subtracted. The middle panel shows the reconstructed source and its position relative to the tangential (astroid) caustic. The left panel shows the resulting constraints on the central exponent of the dark matter halo ($\rho \propto r^{-\gamma}$) and the stellar mass-to-light ratio of the lens galaxy. The dashed contours show the constraints for the same model using the central velocity dispersion measurement from Koopmans & Treu (2003).

witt 1995, and MG1654+134 (Kochanek 1995a). We illustrate them with the Dye & Warren (2003) results for 0047–2808 in Fig. B.72. The mass distribution consists of the lens galaxy and a cuspy dark matter halo, where Fig. B.72 shows the final constraints on the mass-to-light ratio of the stars in the lens galaxy and the exponent of the central dark matter density cusp ($\rho \propto r^{-\gamma}$). The allowed parameter region closely resembles earlier results using either statistical constraints (Fig. B.32) or stellar dynamics (Fig. B.33). In fact, the results using the stellar dynamical constraint from Koopmans & Treu (2003) are superposed on the constraints from the host in Fig. B.72, with the host providing a tighter constraint on the mass distribution than the central velocity dispersion. The one problem with all these models is that they have too few degrees of freedom in their mass distributions by the standards we discussed in §B.4.6. In particular, we know that four-image lenses require both an elliptical lens and an external tidal shear in order to obtain a good fit to the data (e.g. Keeton, Kochanek & Seljak 1997), while none of these models for the extended sources allows for multiple sources of the angular structure in the potential. In fact, the lack of an external shear probably drives the need for dark matter in the 0047–2808 models. Without dark matter, the decay of the stellar quadrupole and the low surface density at the Einstein ring means that the models generate too small a quadrupole moment to fit the data in the absence of a halo. The dark matter solves the problem both through its own ellipticity and the reduction in the necessary shear with a higher surface density near the ring (recall that $\gamma \propto 1 - \langle\kappa\rangle$). Again see the need for a greater focus on the angular structure of the potential.

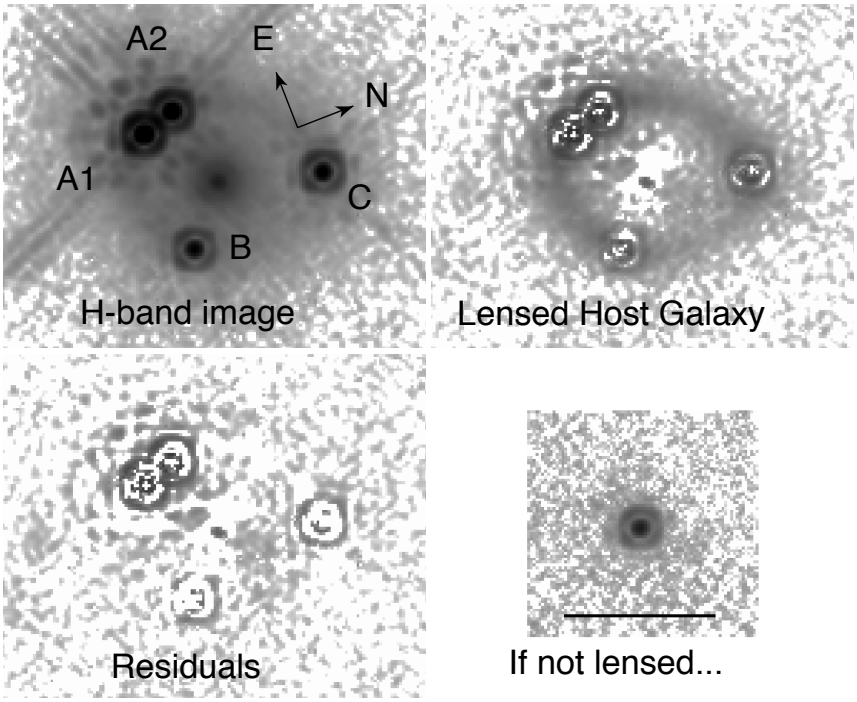


Fig. B.73. The host galaxy in PG1115+080. The top left panel shows the 1-orbit NICMOS image from Impey et al. (1998). The top right panel shows the lensed host galaxy after subtracting the quasar images and the lens galaxy, The lower left panel shows the residuals after subtracting the host as well. For comparison, the lower right panel shows what an image of an unlensed PG1115+080 quasar and host would look like in the same integration time and on the same scales. The host galaxy is an H= 20.8 mag late-type galaxy (Sersic index $n = 1.4$) with a scale length of $R_e = 1.5h^{-1}$ kpc. The demagnified magnitude of the quasar is H= 19.0 mag. The axis ratio of the source, $q_s = 0.65 \pm 0.04$ is consistent with the estimate of $q_s = 0.58 \pm 0.02$ from the simpler ring curve analysis (§B.10.1, Fig. B.71, Kochanek Keeton & McLeod 2001).

B.10.3 Lensed Quasar Host Galaxies

One advantage of studying lensed quasars is that the lens magnification enormously enhances the visibility of the quasar host. A typical HST PSF makes the image of a point source have a mean surface brightness that declines as R^{-3} with distance R from the quasar. Compared to an unlensed quasar, the host galaxy of a lensed quasar is stretched along the Einstein ring leading to an improvement in the contrast between the host in the quasar of μ^2 for an image magnified by μ – you gain μ^3 by stretching the host away from the quasar and lose μ because the quasar is magnified. Perpendicular to the

Einstein ring, the contrast becomes a factor of μ worse than for an unlensed quasar. Since the alignment of the magnification tensor relative to the host changes with each image, the segment of the host where contrast is lost will correspond to a segment where it is gained for another image leading to a net gain for almost all parts of the source when you consider all the images. The distortions produced by lensing also mean the host structure is more easily distinguished from the PSF. In a few cases, like SDSS0924+0219 in Fig. B.58, microlensing or substructure may provide a natural coronograph that supresses the flux from the quasar but not that from the host. Despite naive expectations (and TAC comments), the distortions have little consequence for understanding the structure of the host even though a lens model is required to produce a photometric model of the host.

The only extensive survey of lensed quasar hosts is that of Peng (2004). Fig. B.73 shows the example of PG1115+080, a $z_s = 1.72$ radio-quiet (RQQ) quasar. The Einstein ring image is easily visible even in a short, one-orbit exposure. For comparison, we also took the final model for the quasar and the its host and produced the image that would be obtained in the same time if we observed the quasar in the absence of lensing. It is quite difficult to see the host, and this problem will carry through in any numerical analysis.

At low redshifts ($z < 1$), quasar host galaxies tend to be massive early-type galaxies (e.g. McLure et al. 1999, Dunlop et al. 2003). Over 80% of quasars brighter than $M_V < -23.5$ mag are in early-type galaxies with $L \gtrsim 2L_*$ and effective radii of $R_e \sim 10$ kpc for $z \lesssim 0.5$. Radio quiet quasars (RQQ) tend to be in slightly lower luminosity hosts than radio loud quasars (RLQ) but only by factors of ~ 2 at redshift unity. Far fewer unlensed host galaxies have been detected above redshift unity (e.g. Kukula et al. 2001, Ridgway et al. 2001) with the surprising result that the host galaxies are 2–3 mag brighter than the typical host galaxy at low redshift and corresponded to $\sim 4L_*$ galaxies. Given that the low redshift hosts were already very massive galaxies, it was expected that higher redshift hosts would have lower masses because they were still in the process of being assembled and forming stars (e.g. Kauffmann & Haehnelt 2000). One simple explanation was that by selecting from bright radio sources, these samples picked quasars with more massive black holes as the redshift increased, creating a bias in favor of more massive hosts. The key to checking for such a bias is to be able to detect far less luminous hosts, and the improved surface brightness contrast provided by lensing the host galaxies provides the means.

Fig. B.74 shows the observed H-band magnitudes of the lensed hosts as compared to low redshift host galaxies and other studies of high redshift host galaxies. Although 30% of the lensed quasars are radio-loud, they have luminosities similar to the lensed (or unlensed) radio-quiet hosts. There are no hosts as bright as the Kukula et al. (2001) radio-loud quasar hosts. Once the luminosities of the quasar and the host galaxy are measured we can compare them to the theoretical expectations (Fig. B.75). While the models

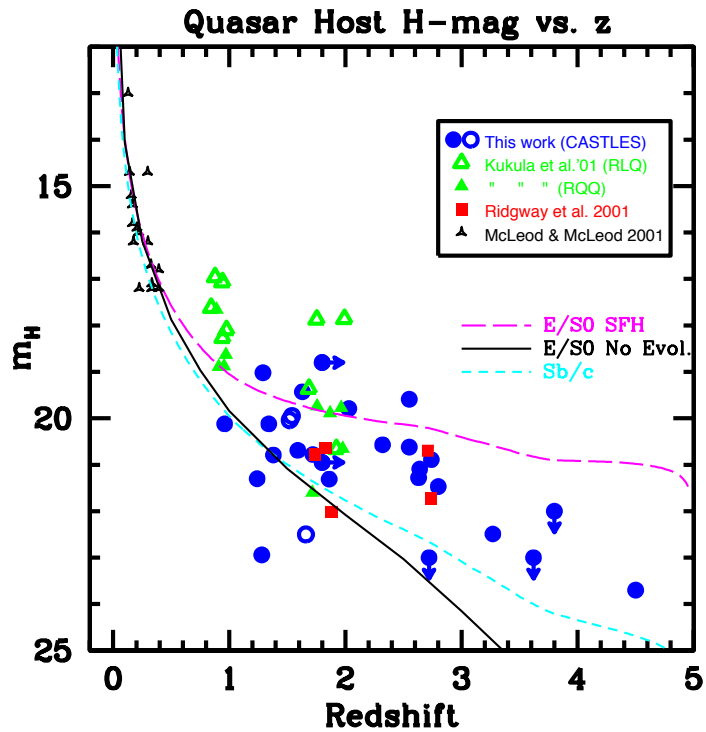


Fig. B.74. Observed H-band magnitudes of quasar host galaxies. The solid (open) circles are secure (more questionable) hosts detected in the CASTLES survey of lensed hosts. The low redshift points are from McLeod & McLeod (2001). All the Ridgway et al. (2001) systems are radio quiet. For comparison, we superpose the evolutionary tracks for a non-evolving E/S0 galaxy (solid curve), an evolving E/S0 galaxy which stars forming stars at $z_f = 5$ with a 1 Gyr exponentially decaying star formation rate (long dashed line) and a star forming Sb/c model (short dashed line). The evolution models are matched to the luminosity of an L_* early-type galaxy at redshift zero. The CASTLES observations can reliably detect hosts about 4 magnitudes fainter than the quasar. From Peng (2004).

agree with the data at low redshift, they are nearly disjoint by $z \sim 3$ in the sense that the observed quasars and hosts are significantly more luminous than predicted. The same holds for the Kukula et al. (2001) and Ridgway et al. (2001) samples, suggesting that black holes masses grow more rapidly than predicted by the theoretical models or that accretion efficiencies were higher in the past. Vestergaard (2004) makes a similar argument based on estimates of black hole masses from emission line widths.

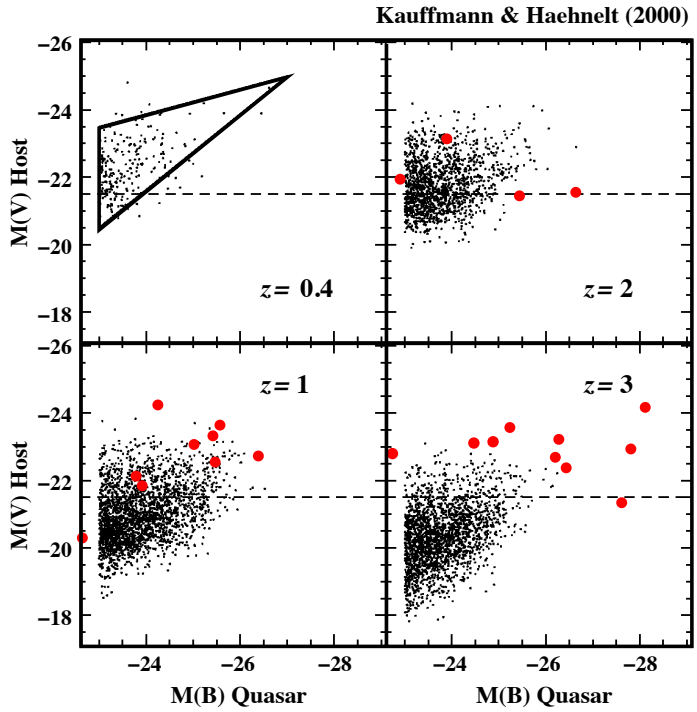


Fig. B.75. A comparison of the estimated rest frame absolute magnitudes of the quasars and hosts as compared to the theoretical models for the evolution of galaxies and the growth of black holes as a function of redshift by Kauffmann & Haehnelt (2000). The low redshift quasars from McLeod & McLeod (2001) occupy the triangle in the upper left panel. At intermediate redshift the lensed host galaxies occupy a region similar to the models, but the two distributions are nearly disjoint by $z \simeq 3$. Both the hosts and the quasars are significantly more luminous than predicted. The horizontal line marks the luminosity of an L_* galaxy at $z = 0$. From Peng (2004).

B.11 Does Strong Lensing Have A Future?

Well, you can hardly expect an answer of "No!" at this point, can you? Since we have just spent nearly 170 pages on the astrophysical uses of lenses, there is no point in reviewing all the results again here. Instead I suggest some goals for the future.

Our first goal is to expand the sample of lenses from ~ 100 to ~ 1000. While 80 lenses seems like a great many compared to even a few years ago, it is still too few to pursue many interesting questions. The problem worsens if the analysis must be limited to lenses meeting other criteria (radio lenses, lenses

found in a well-defined survey, lenses outside the cores of clusters $\cdots$) or if the sample must be subdivided into bins (redshift, separation, luminosity $\cdots$). For example, one of the most interesting applications of lenses will be to map out the halo mass function. This is difficult to do with any other approach because no other selection method works homogeneously on dark low-mass halos, galaxies of different types, groups and clusters. Unlike any other sample in astronomy, gravitational lenses are selected based on mass rather than luminosity, so the same search method works for all halos – the separation distribution of lenses is a direct mapping of the halo mass function. It is not a trivial mapping because the structure of halos changes with mass, but the systematics are far better than those of any other approach. The fact that lenses are mass-selected also gives them an enormous advantage in studying the evolution of galaxies with redshift over optically-selected samples where it will be virtually impossible to select galaxies in the same manner at both low and high redshift. There is no shortage of detectable lenses in the universe – it is simply a question of imaging enough of the sky at high angular resolution. The upgraded VLA and Merlin radio arrays are the most promising tools for this objective.

Our second goal is to systematically monitor the variability of as many lenses as possible. Time delays, if measured in large numbers and measured accurately, can resolve most of the remaining issues about the mass distributions of lenses. This is true even if you regard the H_0 as unmeasured or uncertain – the Hubble constant is the same number for all lenses, so as the number of time delay systems increases, the contribution of the actual value of the Hubble constant to constraining the mass distribution diminishes. At the present time, we are certain that the typical early-type galaxy has a substantial dark matter halo, but we are uncertain how it merges with the luminous galaxy. Steady monitoring of microlensing of the source quasars by the stars in the lens galaxy will also help to resolve this problem because the patterns of the microlensing variability constrain both the stellar surface density near the lensed images and the total density (Part 4, Schechter & Wambsganss 2002). The constraints from time delays and microlensing will be complemented by the continued measurement of central velocity dispersions.

Our third goal should be to obtain ultra-deep, high resolution radio maps of the lenses to search for central images in order to measure the central surface densities of galaxies and to search for supermassive black holes. Keeton (2003a) showed that the dynamic ranges of the existing radio maps of lenses are 1–2 orders of magnitude too small to routinely detect central images given the expected central surface densities of galaxies. Only very asymmetric doubles like PMN1632–0033, where Winn et al. (2004) have detected a central image, are likely to show central images with the present data. Once we reach the sensitivity needed to detect central images, we will also either find central black holes or set strict limits on their existence (Mao, Witt &

Koopmans 2001). This is the only approach that can directly detect even quiescent black holes and determine their masses at cosmological distances. The existing limits could be considerably improved simply by co-adding the existing radio maps either for individual lenses or even for multiple lenses in order to obtain statistical limits.

Our fourth goal should be to unambiguously identify a "dark" satellite of a lens galaxy. For starters we need to conduct complete statistical analyses of lens galaxy satellites in general, by determining their mass functions and radial distributions. As part of such an analysis we can obtain upper bounds on the number of dark satellites. Then, with luck, we will find an example of a lens that requires a satellite at a specific location for which there is no optical counterpart. This may be too conservative a condition. For example, Peng (2004) argues that much of the flux of Object X in MG0414+0534 (Fig. B.6) is actually coming from lensed images of the quasar host galaxy rather than the satellite.

Finally, lens magnification already means that it is far easier to do photometry of a lensed quasar host galaxy than an unlensed galaxy. The next frontier is to measure the kinematics of cosmologically distant host galaxies. This might already be doable for the host galaxy of Q0957+561 at $z_s = 1.41$, but will generally require either JWST or the next generation of ground based telescopes. With larger lens samples we may also find more cases like SDSS0924+0219 where gravitational lensing provides a natural coronograph for the quasar.

Acknowledgments: These lectures are dedicated to Bohdan Paczynski at a very difficult time for a man who made enormous contributions to this field both through his own work and his support for the work of others. I would like to thank E.E. Falco, C.R. Keeton, L.V.E. Koopmans, D. Maoz, J. Muñoz, D. Rusin, D.H. Weinberg and J.N. Winn for commenting on this manuscript, and S. Dye, C.D. Fassnacht, D.R. Marlow, J.L. Mitchell, C.Y. Peng, D. Rusin and J.N Winn for supplying figures. G. Meylan showed considerable patience with the author while waiting for these lectures to be delivered. This research has been supported by the NASA ATP grant NAG5-9265, and by grants HST-GO-7495, 7887, 8175, 8804, and 9133 from the Space Telescope Science Institute. The continuing HST observations of gravitational lenses are an absolutely essential part of converting gravitational lenses from curiosities into astrophysical tools.

References

Agol, E., Jones, B., & Blaes, O., 2000, ApJ, 545, 657
Angonin-Willaime, M.-C., Soucail, G., & Vanderriest, C., 1994, A&A, 291, 411
Bade, N., Siebert, J., Lopez, S., Voges, W., & Reimers, D., 1997, A&A, 317, 13
Barkana, R., 1996, ApJ, 468, 17

Barkana, R., 1997, ApJ, 489, 21
Barkana, R., 1998, ApJ, 502, 531
Barkana, R., Lehár, J., Falco, E.E., Grogin, N.A., Keeton, C.R., & Shapiro, I.I., 1999, ApJ, 523, 54
Bartelmann, M., & Meneghetti, M., 2003, A&A submitted, astro-ph/0312011
Bender, R., et al., 1989, A&A, 217, 35
Bender, R., Saglia, R.P., Ziegler, B., Belloni, P., Greggio, L., Hopp, U., & Bruzual, G., 1998, ApJ, 493, 529
Benson, A.J., Bower, R.G., Frenk, C.S., Lacey, C.G., Baugh, C.M., & Cole, S., 2003, ApJ, 599, 38
Bernstein, G.M., Tyson, J.A., & Kochanek, C.S., 1993, AJ, 105, 816
Bernstein, G., & Fischer, P., 1999, AJ, 118, 14
Beskin, G.M., & Oknyanskij, V.L., 1995, A&A, 304, 341
Biggs, A.D., Browne, I.W.A, Helbig, P., Koopmans, L.V.E., Wilkinson, P.N., & Perley, R.A., 1999, MNRAS, 304, 349
Biggs, A.D., Browne, I.W.A., Jackson, N.J., York, T., Norbury, M.A., McKean, J.P., & Phillips, P.M., 2004, MNRAS submitted [astro-ph/0402128]
Binney, J.J., & Mamon, G.A., 1982, MNRAS, 200, 361
Binney, J.J., & Tremaine, S., 1987, Galactic Dynamics (Princeton: Princeton University Press)
Bischof, O.B., & Becker, R.H., 1997, AJ, 113, 2000
Blandford, R.D., & Narayan, R., 1986, ApJ, 310, 568
Blandford, R.D., & Kochanek, C.S., 1987a, Dark Matter in the Universe, J. Bahcall, T. Piran & S. Weinberg, eds., (World Scientific: Singapore) 133
Blandford, R.D., & Kochanek, C.S., 1987b, ApJ, 321, 658
Blandford, R.D., & Narayan, R., 1992, ARA&A, 30, 311
Blumenthal G.R., Faber S.M., Flores R., & Primack J.R., 1986, ApJ, 301, 27
Bode, P., Ostriker, J.P., & Turok, N., 2001, ApJ, 556, 93
Borgeest, U., von Linde, J., & Refsdal, S., 1991, A&A, 251, L35
Bradac, M., Schneider, P., Steinmetz, M., Lombardi, M., & King, L.J., 2002, A&A, 388, 373
Browne, I.W.A., Wilkinson, P.N., Jackson, N.J.F., et al., 2003, MNRAS, 341, 13
Bruzual, G., & Charlot, S., 1993, ApJ, 405, 538
Bullock, J.S., Kravtsov, A.V., & Weinberg, D.H., 2000, ApJ, 539, 517
Buote, D.A., Jeltema, T.E., Canizares, C.R., & Garmire, G.P., 2002, ApJ,, 577, 183
Burke, B.F., Lehár, J., & Conner, S.R., 1992, inGravitational Lenses, ed.R.Kayser, T.Schramm& L.Nieser (Berlin: Springer), 237
Burkert, A., & Naab, T., 2003, in Galaxies & Chaos, eds. G. Contopulos & N. Voglis (Berlin: Springer Verlag) 327 [astro-ph/0301385]
Burud, I., et al., 2000, ApJ, 544, 117
Burud, I., et al., 2002a, A&A, 383, 71
Burud, I., et al., 2002b, A&A, 391, 481
Byun, Y.-I., Grillmair, C.J., Faber, S.M., Ajhar, E.A., Dressler, A., Kormendy, J., Lauer, T.R., Richstone, D., & Tremaine, S., 1996, AJ, 111, 1889
Cardelli, J.A., Clayton, G.C., & Mathis, J.S., 1989, ApJ, 345, 245
Carroll, S.M., Press, W.H., & Turner, E.L., 1992, ARA&A, 30, 499
Chae, K.-H., Turnshek, D.A., & Khersonsky, V.K., 1998, ApJ, 495, 609
Chae, K.-H., Khersonsky, V.K., & Turnshek, D.A., 1998, ApJ, 506, 80
Chae, K.-H., 1999, ApJ, 524, 582

Chae, K.-H., Mao, S., & Augusto, P., 2001, MNRAS, 326, 1015
Chae, K.H., Biggs, A.D., Blandford, R.D., et al., 2000, PhRvL, 89, 1301
Chae, K.-H., 2003, MNRAS, 346, 746
Chae, K.-H., & Mao, S., 2003, ApJL, 599, L61
Chae, K.-H., Chen, G., Ratra, B., & Lee, D.-W., 2004, ApJ, 607, 71
Chartas, G., Chuss, D., Forman, W., Jones, C., & Shapiro, I., 1998, ApJ, 504, 661
Chartas, G., Bautz, M., Garmire, G., Jones, C., & Schneider, D.P., 2001, ApJL, 550, L163
Chartas, G., Gupta, V., Garmire, G., Jones, C., Falco, E.E., Shapiro, I.I., & Tavecchio, F., 2002, ApJ, 565, 96
Chartas, G., 2003, Carnegie Observatories Astrophysics Series, Vol. 2: Measuring and Modeling the Universe, ed. W. L. Freedman (Pasadena: Carnegie Observatories)
Chen, G.H., & Hewitt, J.N., 1993, AJ, 106, 1719
Chen, G.H., Kochanek, C.S., & Hewitt, J.N., 1995, ApJ, 447, 62
Chen, J., Kravtsov, A.V., & Keeton, C.R., 2003, ApJ, 592, 24
Chiba, M., 2002, ApJ, 565, 17
Claeskens, J., & Surdej, J., 2002, A&ARv 10, 263
Cohen, A.S., Hewitt, J.N., Moore, C.B., & Haarsma, D.B., 2000, ApJ, 545, 578
Cohn, J.D., Kochanek, C.S., McLeod, B.A., & Keeton, C.R., 2001, ApJ, 554, 1216
Cohn, J.D., & Kochanek, C.S., 2004, ApJ in press [astro-ph/0306171]
Cole, S., Lacey, C.G., Baugh, C.M., & Frenk, C.S., 2000, MNRAS, 319, 168
Cole, S., et al., 2001, MNRAS, 326, 255
Courbin, F., Meylan, G., Kneib, J.-P., & Lidman, C., 2002, ApJ, 575, 95
Courbin, F., Saha, P., & Schechter, P., 2002, in Gravitational Lensing: An Astrophysical Tool, F. Courbin, D. Minniti, eds., LRP, 608, 1
Crampton, D., McClure, R.D., & Fletcher, J.M., 1992, ApJ, 392, 23
Dalal, N., & Kochanek, C.S., 2002, ApJ, 572, 25
Dalal, N., & Keeton, C.R., 2003, astro-ph/0312072
Dalal, N., Holz, D.E., Chen, X., & Frieman, J.A., 2003, ApJL, 585, L11
Davis, A.N., Huterer, D., & Krauss, L.M., 2003, MNRAS, 344, 1029
de Vaucouleurs, G., 1948, Ann.Astrophys., 11, 247
Djorgovski, S.G., Perley, R., Meylan, G., & McCarthy, P., 1987, ApJL 321, L17
Djorgovski, S.G., & Davis, M., 1987, ApJ, 313, 59
Djorgovski, S.G., 1991, The Space Distribution of Quasars, ASP #21, D.Crampton, ed., (ASP) 349
Dressler, A., Lynden-Bell, D., Burstein, D., Davies, R.L., Faber, S.M., Terlevich, R.J., & Wegner, G., 1987, ApJ, 313, 42
Dubinski, J., 1992, ApJ, 401, 441
Dubinski, J., 1994, ApJ, 431, 617
Dunlop, J.S., McLure, R.J., Kukula, M.J., Baum, S.A., O'Dea, C.P. & Hughes, D.H., 2003, MNRAS, 340, 1095
Dye, S., & Warren, S., 2003, astro-ph/0311185
Ellithorpe, J.D., Kochanek, C.S., & Hewitt, J.N., 1996, ApJ, 464, 556
Evans, N.W., & Wilkinson, M.I., 1998, MNRAS, 296, 800
Evans, N.W., & Witt, H.J., 2001, MNRAS, 327, 1260
Evans, N.W., & Hunter, C., 2002, ApJ, 575, 68
Evans, N.W., & Witt, H.J., 2003, MNRAS, 345, 1351
Faber, S.M., & Jackson, R.E., 1976, 204, 668

Faber, S.M., Dressler, A., Davies, R.L, Burstein, D., & Lynden-Bell, D., 1987, in Nearly Normal Galaxies, ed. S.M. Faber (New York: Springer) 175
Falco, E.E., Gorenstein, M.V., & Shapiro, I.I., 1985, ApJ, 289, L1
Falco, E.E., Shapiro, I.I., Moustakas, L.A., & Davis, M., 1997, ApJ, 484, 70
Falco, E.E., Kochanek, C.S., & Muñoz, J.A., 1998, ApJ, 494, 47
Falco, E.E., Impey, C.D., Kochanek, C.S., Lehár, J., McLeod, B.A., Rix, H.-W., Keeton, C.R., Muñoz, J.A., & Peng, C.Y., 1999, ApJ, 523, 617
Fassnacht, C.D., Pearson, T.J., Readhead, A.C.S., Browne, I.W.A., Koopmans, L.V.E., Myers, S.T., & Wilkinson, P.N., 1999, ApJ, 527, 498
Fassnacht, C.D., Xanthopoulos, E., Koopmans, L.V.E., & Rusin, D., 2002, ApJ, 581, 823
Fassnacht, C.D., & Lubin, L.M., 2002, AJ, 123, 627
Finch, T., Carlivati, L., Winn, J.N., & Schechter, P.L., 2002, ApJ, 577, 51
Fischer, P., Bernstein, G., Rhee, G., & Tyson, J.A., 1997, AJ, 113, 521
Florentin-Nielsen, R., 1984, A&A, 138, L19
Flores, R., & Primack, J.R., 1996, ApJL, 457, L5
Foltz, C.B,.Hewitt, P.C., Webster, R.L., & Lewis, G.F., 1992, ApJL, 386, L43
Franx, M., Illingworth, G, & de Zeeuw, T., 1991, ApJ 383, 112
Freedman, W., et al., 2001, ApJ, 553, 47
Fukugita, M., Futamase, T., & Kasai, M., 1990, MNRAS, 246, P24
Fukugita, M., & Turner, E.L., 1991, MNRAS, 253, 99
Fukugita, M., Hogan, C.J., & Peebles, P.J.E., 1998, ApJ, 503, 518
Garrett, M.A., Calder, R.J., Porcas, R.W., King, L.J., Walsh, D., & Wilkinson, P.N., 1994, MNRAS, 270, 457
Gebhardt, K., et al., 2003, ApJ in press, astro-ph/0307242
Gerhard, O., Kronawitter, A., Saglia, R.P., & Bender, R., 2001, AJ, 121, 1936
Gerin, M., Phillips, T.G., Benford, D.J., Young, Y.H., Menten, K.M., & Frye, B., 1997, ApJL, 488, 31
Gonzalez, A.H., Williams, K.A., Bullock, J.S., Kolatt, T.S., & Primack, J.R., 2000, ApJ, 528, 145
Gorenstein, M.V., Falco, E.E., & Shapiro, I.I., 1988, ApJ, 327, 693
Gott, J.R., Park, M.-G., & Lee, H.M 1989, ApJ, 338, 1
Green, P.J., Kochanek, C.S., Siemiginowska, A., Kim, D.-W., Markevitch, M., Silverman, J., Dosaj, A., Jannuzi, B.T., & Smith, C., 2002, ApJ, 571, 721
Grogin, N.A., & Narayan, R., 1996, ApJ, 464, 92; erratum, 1996, ApJ, 473, 570
Haarsma, D.B., Hewitt, J.N., Lehár, J., & Burke, B.F., 1999, ApJ, 510, 64
Helbig, P, & Kayser, R., 1996, A&A, 308, 359
Hernquist, L., 1990, ApJ, 356, 359
Hewett, P.C., Foltz, C.B., Harding, M.E., & Lewis, G.F., 1998, AJ, 115, 383
Hewitt, J.N., Turner, E.L., Schneider, D.P., Burke, B.F., & Langston, G.I., 1988, Nature, 333, 537
Hewitt, J.N., Turner, E.L., Lawrence, C.R., Schneider, D.P., & Brody, J.P., 1992, AJ, 104, 968
Heyl, J.S., Hernquist, L., & Spergel, D.N., 1994, ApJ, 427, 165
Hinshaw, G., & Krauss, L.M., 1987, ApJ, 320, 468
Hjorth, J., et al., 2002, ApJ, 572, L11
Holder, G.P., & Schechter, P.L., 2003, ApJ, 589, 688
Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E., & Perley, R., 1985, AJ, 90, 691

Hunter, C., & Evans, N.W., 2001, ApJ, 554, 1227
Huterer, D., & Ma, C.-P., 2004, ApJ, 600, 7
Ibata, R.A., Lewis, G.F., Irwin, M.J., Lehár, J., & Totten, E.J., 1999, AJ, 118, 1922
Im, M., Griffiths, R.E., & Ratnatunga, K.U., 1997, ApJ, 475, 457
Impey, C.D., Falco, E.E., Kochanek, C.S., Lehár, J., McLeod, B.A., Rix, H.-W., Peng, C.Y., & Keeton, C.R., 1998, ApJ, 509, 551
Inada, N., Oguri, M., Pindor, B., et al., 2003, Nature, 426, 810
Irwin, M.J., Ibata, R.A., Lewis, G.F., & Totten, E.J., 1998, ApJ, 505, 529
Jaffe, W., 1983, MNRAS, 202, 995
Jaunsen, A.O., & Hjorth, J., 1997, A&Ap, 317, L39
Jean, C., & Surdej, J., 1998, A&Ap, 339, 729
Johnston, D.E., Richards, G.T., Frieman , J.A., et al., 2003, AJ, 126, 2281
Jones, D.L., et al., 1996, ApJ, 470, 23
Jorgensen, I., Franx, M., & Kjaergaard, P., 1996, MNRAS, 280, 167
Kassiola, A., & Kovner, I., 1993, ApJ, 417, 450
Kauffmann, G., White, S.D.M., & Guiderdoni, B., 1993, MNRAS, 264, 201
Kauffmann, G., & Haehnelt, M., 2000, MNRAS, 311, 576
Kawano, Y., Oguri, M., Matsubara, T., & Ikeuchi, S., 2004, PASJ in press, astro-ph/0404013
Kazantzidis, S., Kravtsov, A.V., Zentner, A.R., Algood, B., Nagai, D., & Moore, B., 2004, ApJL submitted [astro-ph/0405189]
Keeton, C.R., Kochanek, C.S., & Seljak, U., 1997, ApJ, 482, 604
Keeton, C.R., & Kochanek, C.S., 1998, ApJ, 495, 157
Keeton, C.R., Kochanek, C.S., & Falco, E.E., 1998, ApJ, 509, 561
Keeton, C.R., Falco, E.E., Impey, C.D., Kochanek, C.S., Lehár, J., McLeod, B.A., Rix, H.-W., Muñoz, J.A., & Peng, C.Y., 2000, ApJ, 542, 74
Keeton, C.R., Christlein, D., & Zabludoff, A.I., 2000, ApJ, 545, 129
Keeton, C.R., 2001a, astro-ph/0102340
Keeton, C.R., 2001b, ApJ, 46, 60
Keeton, C.R., & Madau, P., 2001, ApJ, 549, 25
Keeton, C.R., 2002, ApJL, 575, L1
Keeton, C.R., Gaudi, B.S., & Petters, A.O., 2003, ApJ, 598, 138
Keeton, C.R., & Winn, J.N., 2003, ApJ, 590, 39
Keeton, C.R., 2003a, ApJ, 582, 17
Keeton, C.R., 2003b, ApJ, 584, 664
Keeton, C.R., Kuhlen, M., & Haiman, Z., 2004, ApJ submitted [astro-ph/0405143]
Kelson, D.D., van Dokkum, P, Franx, M., Illingworth, G., & Fabricant, D., 1997, ApJ, 478, L13
Kelson, D.D., Illingworth, G.D., van Dokkum, P, & Franx, M., 2000, ApJ, 531, 184
Khairul, A.S.M., & Ryden, B.S., 2002, ApJ, 570, 610
King, L., & Browne, I.W.A., 1996, MNRAS, 282, 67
Klypin, A., Kravtsov, A.V., Valenzuela, O., & Prada, F., 1999, ApJ, 522, 82
Kneib, J.-P., Ellis, R.S., Smail, I., Couch, W.J., & Sharples, R.M., 1996, ApJ, 471, 643
Kneib, J.-P., Cohen, J.G & Hjorth, J., 2000, ApJL, 544, 35
Kochanek, C.S., & Blandford, R.D., 1987, ApJ, 321, 676
Kochanek, C.S., & Apostolakis, J., 1988, MNRAS, 235, 1073
Kochanek, C.S., Blandford, R.D., Lawrence, C.R., & Narayan, R., 1989, MNRAS, 238, 43

Kochanek, C.S., & Lawrence, C.R., 1990, AJ, 99, 1700
Kochanek, C.S., 1991a, ApJ, 373, 354
Kochanek, C.S., 1991b, ApJ, 379, 517
Kochanek, C.S., 1992a, ApJ, 384, 1
Kochanek, C.S., 1992b, ApJ, 397, 381
Kochanek, C.S., & Narayan, R., 1992, ApJ, 401, 461
Kochanek, C.S., 1993a, ApJ, 417, 438
Kochanek, C.S., 1993b, ApJ, 419, 12
Kochanek, C.S., 1993c, MNRAS, 261, 453
Kochanek, C.S., 1994, ApJ, 436, 56
Kochanek, C.S., 1995a, ApJ, 445, 559
Kochanek, C.S., 1995b, ApJ, 453, 545
Kochanek, C.S., Falco, E.E., & Schild, R., 1995, ApJ 452, 109
Kochanek, C.S., 1996a, ApJ, 466, 638
Kochanek, C.S., 1996b, ApJ, 473, 595
Kochanek, C.S., Falco, E.E., & Muñoz, J.A., 1999, ApJ, 510, 590
Kochanek, C.S., et al., 2000, ApJ, 543, 131
Kochanek, C.S., Falco, E.E., Impey, C.D., Lehár, J., McLeod, B.A., Rix, H.-W., Keeton, C.R., Peng, C.Y., & Muñoz, J.A., 2000, ApJ, 535, 692
Kochanek, C.S., Keeton, C.R., & McLeod, B.A., 2001, ApJ, 547, 50
Kochanek, C.S., & White, M., 2001, ApJ 559, 531
Kochanek, C.S., Pahre, M.A., Falco, E.E., Huchra, J.P., Mader, J., Jarrett, T.H., Chester, T., Cutri, R., & Schneider, S.E., 2001, ApJ, 560, 566
Kochanek, C.S., 2002a, ApJ, 578, 25
Kochanek, C.S., 2002b, in Shapes of Galaxies and Their Dark Halos, P., Natarajan, ed., (Singapore: World Scientific) 62
Kochanek, C.S., 2003a, ApJ, 583, 49
Kochanek, C.S., 2003b, astro-ph/0204043
Kochanek, C.S., 2003c, in The Dark Universe: Matter Energy and Gravity, M. Livio, ed. (Cambridge Univ. Press: Cambridge) 139
Kochanek, C.S., & Schechter, P., 2004, in Measuring and Modeling the Universe (Carnegie Observatories Centennial Symposia) ed. W.L Freedman (Cambridge: Cambridge Univ. Press) 117
Kochanek, C.S., 2004, ApJ, 605, 58
Kochanek, C.S., & Dalal, N., 2004, ApJ in press, astro-ph/0302036
Koopmans, L.V.E., de Bruyn, A.G., & Jackson, N., 1998, MNRAS, 295, 534
Koopmans, L.V.E., de Bruyn, A.G., Xanthopoulos, E., & Fassnacht, C.D., 2000, A&A, 356, 391
Koopmans, L.V.E., de Bruyn, A.G., Fassnacht, C.D., Marlow, D.R., et al., 2000, A&A, 361, 815
Koopmans, L.V.E., Garrett, M.A., Blandford, R.D., et al., 2002, MNRAS, 334, 39
Koopmans, L.V.E., & Treu, T., 2002, ApJ, 568, 5
Koopmans, L.V.E., & Treu, T., 2003, ApJ, 583, 606
Koopmans, L.V.E., Treu, T., Fassnacht, C.D., Blandford, R.D., & Surpi, G., 2003, ApJ, 599, 70
Koopmans, L.V.E., & de Bruyn, A.G., 2003, ApJL submitted [astro-ph/0311567]
Kormann, R., Schneider, P., & Bartelmann, M., 1994, A&A, 284, 285
Kovner, I., 1987a, ApJ, 312, 22
Kovner, I., 1987b, ApJ 316, 52

Krauss, L.M., & White, M., 1992, ApJ, 394, 385
Kuhlen, M., Keeton, C.R., & Madau, P., 2004, ApJ, 601, 104
Kukula, M.J., Dunlop, J.S., McLure, R.J, Miller, L., Percival, W.J., Baum, S.A., & O'Dea, C.P., 2001, MNRAS, 326, 1533
Kundić, T., Turner, E.L., Colley, W.N., Gott, J.R., Rhoads, J.E., Wang, Y., Bergeron, L.E., Gloria, K.A., Long, D.C., Malhotra, S., & Wambsganss, J., 1997, ApJ, 482, 75
Kuzmin, G., 1956, Astr.Zh., 33, 27
Lawrence, C.R., Schneider, D.P., Schmidt, M., Bennett, C.L., Hewitt, J.N., Burke, B.F., Turner, E.L., & Gunn, J.E., 1984, Science, 223, 46
Lehár, J., Cooke, AJ, Lawrence, C.R., Silber, A.D., & Langston, G.I., 1996, AJ, 111, 1812
Lehár, J., Falco, E.E., Kochanek, C.S., McLeod, B.A, Muñoz, J.A., Impey, C.D., Rix, H.-W., Keeton, C.R., & Peng, C.Y., 2000, ApJ, 536, 584
Lehár, J., Buchalter, A., McMahon, R.G., Kochanek, C.S., & Muxlow, T.W.B., 2001, ApJ, 547, 60
Li, L.-X., & Ostriker, J.P., 2002, ApJ, 566, 652
Li, L.-X., & Ostriker, J.P., 2003, ApJ, 595, 603
Lidman, C., Courbin, F., Kneib, J.-P., Golse, G., Castandar, F., & Soucail, G., 2000, A&A, 364, L62
Loewenstein, M., & Mushotzky, R.F., 2003, ApJ, submitted ??ASTROPH
Lovell, J.E.J., Jauncey, D.L., Reynolds, J.E., Wieringa, M.H., King, E.A., Tzioumis, A., McCulloch, P.M., & Edwards, P.G., 1998, ApJ, 508, L51
Ma, C.-P., 2003, ApJL, 584, L1
Maller, A.H., Flores, R.A., & Primack, J.R., 1997, ApJ, 486, 681
Maller, A.H., Simard, L., Guhathakurta, P., Hjorth, J., Jaunsen, A.O., Flores, R.A., & Primack, J.R., 2000, ApJ, 533, 194
Mao, S.D., 1991, ApJ, 380, 9
Mao, S.D., & Kochanek, C.S., 1994, MNRAS, 268, 569
Mao, S.D., & Schneider, P., 1998, MNRAS, 295, 587
Mao, S., Witt, H.J., & Koopmans, L.V.E., MNRAS, 323, 301
Mao, S., Jing, Y., Ostriker, J.P., & Weller, J., 2004, ApJL, 604, 5L
Maoz, D., Bahcall, J.N., Schneider, D.P., et al., 1993, ApJ, 409, 28
Maoz, D., & Rix, H.-W., 1993, ApJ, 416, 425
Maoz, D., Rix, H.-W., Gal-Yam, A., & Gould, A., 1997, ApJ, 486, 75
Marlow, D.R., Browne, I.W.A., Jackson, N., & Wilkinson, P.N., 1999, MNRAS, 305, 15
Marlow, D.R., Rusin,¡ D., Jackson, N., Wilkinson, P.N., Browne, I.W.A., & Koopmans, L., 2000, AJ, 119, 2629
McKay, T.A., et al., 2002, ApJ, 571, L85
McLeod, K.K., & McLeod, B.A., 2001, ApJ, 546, 782
McLure, R.J., Kukula, M.J., Dunlop, J.S., Baum, S.A., O'Dea, C.P., & Hughes, D.H., 1999, MNRAS, 308, 377
Mestel, L., 1963, MNRAS, 126, 553
Metcalf, R.B., & Madau, P., 2001, ApJ, 563, 9
Metcalf, R.B., 2002, ApJ, 580, 696
Metcalf, R.B., & Zhao, H., 2002, ApJ, 567, L5
Miralda-Escude, J., & Lehár, J., 1992, MNRAS, 259, 31p
Mitchell, J.L., Keeton, C.R., Frieman, J.A., & Sheth, R.K., 2004, ApJ submitted [astro-ph/0401138]

Mo, H.J., Mao, S., & White, S.D.M., 1998, MNRAS, 295, 319
Möller, O., Hewett, P., & Blain, A.W., 2003, MNRAS, 345, 1
Möller, O., & Blain, A.W., 2001, MNRAS, 327, 339
Moore, B., et al., 1998, ApJ, 499, L5
Moore, B., 1999, MNRAS, 310, 1147
Moore, B.Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., & Tozzi, P., 1999, ApJ, 524, L19
Morgan, N.D., Chartas, G., Malm, M., Bautz, M.W., Burud, I., Hjorth, J., Jones, S.E., & Schechter, P.L., 2001, ApJ, 555, 1
Morgan, N.D., Snyder, J.A., & Reens, L.H., 2003, AJ, 126, 2145
Mortlock, D.J., Webster, R.L., & Francis, P.J., 1999, MNRAS, 309, 836
Mortlock, D.J., & Webster, R.L., 2000a, MNRAS, 319, 860
Mortlock, D.J., & Webster, R.L., 2000b, MNRAS, 319, 872
Mortlock, D.J., & Webster, R.L., 2000c, MNRAS, 319, 879
Mortlock, D.J., & Webster, R.L., 2001, MNRAS, 321, 629
Motta, V., Mediavilla, E., Muñoz, J.A., Falco, E.E., et al., 2002, ApJ, 574, 719
Moustakas, L.A., & Metcalf, R.B., 2003, MNRAS, 339, 607
Muñoz, J.A., Falco, E.E., Kochanek, C.S., Lehár, J., Herold, L.K., Fletcher, A.B., & Burke, B.F., 1998, ApJL, 492, L9
Muñoz, J.A., Kochanek, C.S., & Falco, E.E., 1999, ApJL, 521, L17
Muñoz, J.A., Kochanek, C.S., & Keeton, C.R., 2001, ApJ, 558, 657
Muñoz, J.A., Falco, E.E., Kochanek, C.S., Lehár, J., & Mediavilla, E., 2003, ApJ, 594, 684
Muñoz, J.A., Falco, E.E., Kochanek, C.S., McLeod, B.A., & Mediavilla, E., 2004, ApJ, 605, 614
Myers, S.T., Fassnacht, C.D., Djorgovski, S.G., et al., 1995, ApJL, 447, L5
Myers, S.T., Rusin, D., Fassnacht, C.D., Blandford, R.D., Pearson, T.J., Readhead, A.C.S., Jackson, N., Browne, I.W.A., Marlow, D.R., Wilkinson, P.N., Koopmans, L.V.E., & de Bruyn, A.G., 1999, AJ, 117, 2565
Nadeau, D., Yee, H.K.C., Forrest, W.J., Garnett, J.D., Ninkov, Z., & Pipher, J.L., 1991, ApJ, 376, 430
Narayan, R., & White, S.D.M., 1988, MNRAS, 231, P97
Narayan, R., & Bartelmann, M., in Formation of Structure in the Universe, A. Dekel & J. Ostriker, eds., (Cambridge: Cambridge University Press) 360
Navarro, J., Frenk, C.S., & White, S.D.M., 1996, ApJ, 462, 563
Ofek, E.O., Maoz, D., Prada, F., Kolatt, T., & Rix, H.-W., 2001, MNRAS, 324, 463
Ofek, E.O., & Maoz, D., 2003, ApJ, 594, 101
Ofek, E.O., Rix, H.-W., & Maoz, D., 2003, MNRAS, 343, 639
Oguri, M., & Keeton, C.R., 2004, ApJ submitted, astro-ph/0403633
Oguri, M., Inada, N., Keeton, C.R., Pindor, B., Hennawi, J.F., Gregg, M.D., Becker, R.H., et al., 2004, ApJ, 604, 78
Ohyama, Y., et al., 2002, AJ, 123, 2903
Olling, R.P., & Merrifield, M., 2001, MNRAS, 311, 361
Pahre, M.A., de Carvalho, R.R., & Djorgovski, S.G., 1998, AJ, 116, 1606
Patnaik, A.R., & Narasimha, D., 2001, MNRAS, 326, 1403
Pearce, F.R., Jenkins, A., Frenk, C.S., Colberg, J.M., White, S.D.M., Thomas, P.A., Couchman, H.M.P., Peacock, J.A., & Efstathiou, G., 1999, ApJ, 521, L99
Pello, R., Miralles, J.M., Le Borgne, J.-F., Picat, J.-P., Soucail, G.., & Bruzual, G., 1996, A&A, 314, 73

Pelt, J., Kayser, R., Refsdal, S., & Schramm, T., 1996, A&A, 305, 97
Peng, C.Y., 2004, PhD Thesis, Steward Observatory, University of Arizona
Phillips, P.M., Browne, I.W.A., Jackson, N.J., Wilkinson, P.N., Mao, S., Rusin, D., Marlow, D.R., Snellen, I., & Nesser, M., 2001, MNRAS, 328, 1001
Pindor, B., Turner, E.L., Lupton, R.H., & Brinkmann, J., 2003, AJ, 125, 2325
Porciani, C., & Madau, P., 2000, ApJ, 532, 679
Press, W.H., & Gunn, J.E., 1973, ApJ, 185, 397
Press, W.H., & Schechter, P., 1974, ApJ, 187, 425
Press, W.H., Rybicki, G.B., & Hewitt, J.N., 1992a, ApJ, 385, 404
Press, W.H., Rybicki, G.B., & Hewitt, J.N., 1992b, ApJ, 385, 416
Quadri, R., Möeller, O., & Natarajan, P., 2003, ApJ, 597, 659
Ratnatunga, K.U., Griffiths, R.E., & Ostrander, E.J., 1999, AJ, 117, 2010
Rees, M.J., & Ostriker, J.P., 1977, MNRAS, 179, 541
Refsdal, S., 1964, 128, 307
Refsdal, S., & Surdej, J., 1994, RPPh, 57, 117
Rest, A., van den Bosch, F.C., Jaffe, W., Tran, H., Tsvetanov, Z., Ford, H.C., Davies, J., & Schafer, J.L, 2001, AJ, 121, 2431
Ridgway, S.E., Heckman, T.M., Calzetti, D., & Lehnert, M., 2001, ApJ, 550, 122
Riess, A.G., Strolger, L.-G., Tonry, J., Casertano, S., et al., 2004, ApJ, 607, 665
Rix, H.-W., Maoz, D., Turner, E.L., & Fukugita, M., 1994, ApJ, 435, 49
Romanowsky, A.J., & Kochanek, C.S., 1998, ApJ, 493, 64
Romanowsky, A.J., & Kochanek, C.S., 1999, ApJ, 516, 18
Ros, E., Guirado, J.C., Marcaide, J.M., Perez-Torres, M.A., Falco, E.E., Muñoz, J.A., Alberdi, A., & Lara, L., 2000, A&A, 362, 845
Rusin, D., & Ma, C.-P., 2001, ApJ, 549, L33
Rusin, D., Kochanek, C.S., Norbury, M., et al., 2001, ApJ, 557, 594
Rusin, D., & Tegmark, M., 2001, ApJ, 553, 709
Rusin, D., et al., 2002, MNRAS, 330, 205
Rusin, D., 2002, ApJ, 572, 705
Rusin, D., Kochanek, C.S., Falco, E.E., Keeton, C.R., McLeod, B.A., Impey, C.D., Lehár, J., Munoz, J.A., Peng, C.Y., & Rix, H.-W., 2003, ApJ, 587, 143
Rusin, D., Kochanek, C.S., & Keeton, C.R., 2003, ApJ, 595, 29
Rusin, D., & Kochanek, C.S., 2004, ApJ submitted
Saha, P., & Williams, L.R.L., 1997, MNRAS, 292, 148
Saha, P., 2000, AJ, 120, 1654
Saha, P., & Williams, L.L.R., 2001, AJ, 122, 585
Saha, P., 2004, A&A, 414, 425
Saha, P., & Williams, L.L.R., 2004, AJ in press, astro-ph/0402135
Sand, D.J., Treu, T., & Ellis, R.S., 2002, ApJL, 574, L129
Sand, D.J., Treu, T., Smith, G.P., & Ellis, R.S., 2003, astro-ph/0300703
Schechter, P., 1976, ApJ 203, 297
Schechter, P.L., & Moore, C.B., 1993, AJ, 105, 1
Schechter, P.L., Bailyn, C.D., Barr, R., et al., 1997, ApJL, 475, L85
Schechter, P.L., & Wambsganss, J., 2002, ApJ, 580, 685
Schechter, P.L., Udalski, A., Szymanski, M., Kubiak, M., Pietrzynski, G., Soszynski, I., Wozniak, P., Zebrun, K., Szewczyk, O., & Wyrzykowki, L., 2003, ApJ, 584, 657
Schild, R.E., & Cholfin, B., 1986, ApJ, 300, 209
Schild, R.E., & Thomson, D.J., 1995, AJ, 109, 1970

Schild, R.E., & Thomson, D.J., 1997, AJ, 113, 130
Schneider, P., Ehlers, J., & Falco, E.E., 1992, Gravitational Lenses, (Berlin: Springer-Verlag)
Scodeggio, M., Giovanelli, R., & Haynes, M.P., 1998, ApJ, 116, 2738
Sersic, J.-L., 1968, Atlas de Galaxias Australes (Cordoba: Obs. Astron.)
Sheth, R.K., & Tormen, G., 1999, MNRAS 308, 119
Sheth, R.K., Bernardi, M., Schechter, P.L., et al., 2003, ApJ, 594, 225
Sluse, D., Surdej, J., Claeskens, J.-F., Hutsemekers, D., Jean, C., Courbin, F., Nakos, T., Billeres, M., & Khmil, S.V., 2003, A&A, L43
Small, T.A., Sargent, W.L.W., & Steidel, C.C., 1997, AJ, 114, 2254
Soucail, G., Kneib, J.-P., & Golse, G., 2004, A&A, 417, 33
Spergel, D., & Steinhardt, P., 2000, PRL, 84, 3760
Spergel, D.N., Verde, L., Peiris, H.V., et al., 2003, ApJS, 148, 175
Surdej, J., Claeskens, J.F., Crampton, D., et al., 1993, AJ, 105, 2064
Sykes, C.M., Browne, I.W.A., Jackson, N.J., Marlow, D.R., Nair, S., Wilkinson, P.N., Blandford, R.D., Cohen, J., Fassnacht, C.D., Hogg, D., Pearson, T.J., Readhead, A.C.S., Womble, D.S., Myers, S.T., de Bruyn, A.G., Bremer, M., Miley, G.K., & Schilizzi, R.T., 1998, MNRAS, 301, 310
Toft, S., Hjorth, J., Burud, I., 2000, A&A, 357, 115
Tonry, J.L., 1998, AJ, 115, 1
Tonry, J.L., & Franx, M., 1999, ApJ, 515, 512
Treu, T., Stiavelli, M., Bertin, G., Casertano, S., & Moller, P., 2001, MNRAS, 326, 327
Treu, T., Stiavelli, M., Casertano, S., Moller, P., & Bertin, G., 2002, ApJL, 564, L13
Treu, T., & Koopmans, L.V.E., 2002a, MNRAS, 337, P6
Treu, T., & Koopmans, L.V.E., 2002b, ApJ, 575, 87
Treu, T., & Koopmans, L.V.E., 2004, ApJ submitted, astro-ph/0401373
Trott, C.M., & Webster, R.L., 2002, MNRAS, 334, 621
Trotter, C.S., Winn, J.N., & Hewitt, J.N., 2000, ApJ, 535, 671
Tully, R.B., & Fisher, B., 1977, A&A, 54, 661
Turner, C.M., Keeton, C.R., & Kochanek, C.S., 2004, in preparation
Turner, E.L., 1980, ApJL, 242, L135
Turner, E.L., Ostriker, J.P., & Gott J.R., 1984, ApJ, 284, 1
Vanderriest, C., Schneider, J., Herpe, G., Chevreton, M., Moles, M., & Wlerick, G., 1989, A&A, 215, 1
van de Ven, G., van Dokkum, P.G., & Franx, M., 2003, MNRAS, 344, 924
van Dokkum, P.G., & Franx, 1996, MNRAS, 281, 985 & Kelson, D.D., 2000, ApJL, 541, 95
van Dokkum, P.G., & Franx, M., 2001, ApJ, 553, 90
van Dokkum, P.G., Franx, Kelson, D.D., & Illingworth, G.D., 2001, ApJL, 553, L39
van Dokkum, P.G., & Ellis, R.S., 2003, ApJL, 592, L53
van der Marel, R.P., & Franx, M., 1993, ApJ, 407, 525
Vestergaard, M., 2004, ApJ, 601, 676
Wallington, S., & Narayan, R., 1993, ApJ, 403, 517
Wallington, S., Kochanek, C.S., & Koo, D.C., 1995, ApJ, 441, 58
Wallington, S., Kochanek, C.S., & Narayan, R., 1996, ApJ, 465, 64
Walsh, D., Carswell, R.F., & Weymann, R.J., 1979, Nature, 279, 381
Wambsganss, J., & Paczynski, B., 1994, AJ, 108, 1156

Wambsganss, J., Cen, R., Ostriker, J.P., & Turner, E.L., 1995, Science, 268, 274
Wambsganss, J., Cen, R., & Ostriker, J.P., 1998, ApJ, 494, 29
Wambsganss, J., 1998, LRR, 1, 12
Warren, M.S., Quinn, P.J., Salmon, J.K., & Zurek, W.H., 1992, ApJ, 399, 405
Warren, S.J., Hewett, P.C., Lewis, G.F., Moller, P., Iovino, A., & Shaver, P.A., 1996, MNRAS, 278, 139
Warren, S.J., & Dye, S., 2003, ApJ, 590, 673
Weedman, D.W., Weymann, R.J., Green, R.F., & Heckman, T.M., 1982, ApJL, 255, L5
Weymann, R.J., Latham, D., Roger, J., Angel, P., Green, R.F., Liebert, J., Turnshek, D.A., Turnshek, D.E., & Tyson, J.A., 1980, Nature, 285, 641
White, M., Hernquist, L., & Springel, V., 2001, astro-ph/0107023
Wiklind, T., & Combes, F., 1996, Nature, 379, 139
Wiklind, T., & Alloin, D., 2002, in Gravitational Lensing: An Astrophysical Tool, F. Courbin & D. Minniti, eds., Lecture Notes in Physics, 608, 124
Wilkinson, P.N., Henstock, D.R., Browne, I.W.A., Readhead, A.C.S., Pearson, T.J., Taylor, G.B., & Vermeulen, R.C., 2001, in Gravitational Lensing: Recent Progress Future Goals (ASP #237), T.G. Brainerd & C.S. Kochanek, eds., (San Francisco: ASP0 37
Williams, L.L.R., & Saha, P., 2000, AJ, 119, 439
Winn, J.N., Hewitt, J.N., & Schechter, P.L., 2001, Gravitational Lensing: Recent Progress and Future Goals, eds., T.G. Brainerd & C.S. Kochanek (San Francisco: ASP) 61
Winn, J.N., et al., 2002a, AJ, 123, 10
Winn, J.N., Kochanek, C.S., McLeod, B.A., Falco, E.E., Impey, C.D., & Rix, H.-W., 2002b, ApJ, 575, 103
Winn, J.N., Lovell, J.E.J., Chen, H., Fletcher, A., Hewitt, J.N., Patnaik, A.R., & Schechter, P.L., 2002c, ApJ, 564, 143
Winn, J.N., Rusin, D., & Kochanek, C.S., 2003, ApJ, 587, 80
Winn, J.N., Kochanek, C.S., Keeton, C.R., & Lovell, J.E.J., 2003, ApJ, 590, 26
Winn, J.N., Hall, P.B., & Schechter, P.L., 2003, ApJ, 597, 672
Winn, J.N., Rusin, D.S., & Kochanek, C.S., 2004, Nature, 427, 613
Wisotzki, L., Kohler, T., Kayser, R., & Reimers, D., 1993, A&A, 278, L15
Wisotzki, L., Schechter, P.L., Bradt, H.V., Heinmuller, J., & Reimers, D., 2002, A&A, 395, 17
Witt, A.N.., Thronson, H.A., & Capuano, J.M., 1992, ApJ, 393, 611
Witt, H.J., & Mao, S., 2000, MNRAS, 311, 689
Witt, H.J, Mao, S., & Keeton, C.R., 2000, ApJ, 544, 98
Wucknitz, O., 2002, MNRAS, 332, 951
Wucknitz, O., 2004, MNRAS, 349, 1
Wucknitz, O., Biggs, A.D., & Browne, I.W.A., 2004, MNRAS, 349, 14
Wyithe, J.S.B., Turner, E.L., & Spergel, D.N., 2001, ApJ, 555, 504
Wyithe, J.S.B., Winn, J.N., & Rusin, D., 2003, ApJ, 583, 58
Wyithe, J.S.B., 2004, MNRAS in press
Wyrzykowski, L., Udalski, A., Schechter, P.L., Szewczyk, O., Szymanski, M., Kubiak, M., Piertrzynski, G., Soszynski, I., & Zebrun, K., 2003, AcA, 53, 229
Yee, H.K.C., Fillipenko, A.V., & Tang, D., 1993, AJ, 105, 7
York, T., Jackson, N., Browne, I.W.A., Wucknitz, O., & Skelton, J.E., 2004, MNRAS submitted [astro-ph/0405115]

Zaritsky, D., & Rix, H.-W., 1997, AJ, 477, 118
Zentner, A.R., & Bullock, J.S., 2003, ApJ, 598, 49
Zhao, H., 1997, MNRAS, 287, 525
Zhao, H., & Pronk, D., 2001, MNRAS, 320, 401