Figure 5.1: Sizes of the Einstein radii of SIS lenses of different masses (velocity dispersions) as a function of the lens redshift and for different source redshifts.

One can then imagine each component in this sum as the superposition of several surface density modes corresponding to different spatial scales. Following this idea, we can write the surface density as a multipole expansion:

$$\kappa(\vec{x}) = \kappa_0(x) + \sum_{m=1}^{\infty} \kappa_m(x) \exp(im\phi) \ . \tag{5.5}$$

Note that in principle one single multiple expansion could be sufficient to describe the whole lens mass distribution. However, most implementation of parametric modeling use the parametric form in Eq. 5.4 and use a limited set of multipoles to describe each component (smooth or clumpy);

- in this expansion the term κ_0 is the monopole and it describes the axially symmetric part of the lens; the term with $m = 1$ is the dipole; the term with $m = 2$ is the quadrupole, which gives the degree of **ellipticity** of the iso-density contours. If a "power spectrum" is calculated with the coefficients κ_m, these result to be the most important terms. Higher order terms of the multipole expansion describe structures on increasingly smaller scales;

- the **density profile** $\kappa(x)$ $[\rho(r)]$ is a fundamental property of the lens. Differ-

ent profiles produce remarkably different lensing features, for example different multiplicities of the images, different strengths of the shear field and of the magnification pattern on the source and on the lens planes. We will characterize the profile in terms of a set of parameters, $\mathbf{q}$: for example, in the case of the SIS, we will just need one parameter (the velocity dispersion) to uniquely identify the profile of the lens. If we adopt a NIS profile, we will need one additional parameter, namely the core size;

- If the amount to substructures to be included in the model is too large compared to the number of lensing constraints available, we cannot afford to model each of them independently. To reduce dramatically the number of parameters brought into the model by the substructures, one can assume that these must obey to some **scaling relations**. This solution is widely used to model the effects of the cluster galaxies, which are generally assumed to trace the cluster substructures. Let us assume that galaxies are modeled with isothermal profiles. Their velocity dispersion may be linked to the galaxy luminosity using the Faber-Jackson relation, i.e. we may assume that a relation of the kind

$$ \sigma = \sigma_\star \left(\frac{L}{L_\star} \right)^{1/4} \tag{5.6} $$

holds. Since galaxies (or more generally speaking, substructures) are tidally stripped of their external halo, an additional parameter (the truncation radius) is often needed to truncate the mass profile. Even in this case, we can use a global scaling relation to bind the parameter to the observable (the luminosity):

$$ r_t = r_{t,\star} \left(\frac{L}{L_\star} \right)^{\eta} . \tag{5.7} $$

Following this method, we describe the whole cluster galaxy population with only few parameters: $\mathbf{s} = (L_\star, \sigma_\star, r_{t,\star}, \eta)$

- Additional terms in the multipole expansion will bring into the model additional parameters, $\mathbf{m}$. Likely, we will use at least elliptical mass distributions to describe the smooth and clumpy components, which are defined by the **axis ratios** f_i and by the **position angle** φ_i;

- Finally, each component will also need to be centered properly, typically following the light distribution, but sometime leaving some degree of freedom. The centers themselves are parameters defining the lens model, $\mathbf{x_c} = [\mathbf{x_{c,i}}]$.

Summarizing, the mass distribution of the lens is fully defined by a set of parameters, $\mathbf{p} = (\mathbf{q}, \mathbf{m}, \mathbf{s}, \mathbf{x_c})$. This modeling approach is thus called *parametric*. As it appears clear from the recipe we just outline, it somehow relies on the assumption that the light traces the mass. This is may be the strongest limitation of this approach.

5.2.3 Observables

What can we learn from images like that in Fig. 5.2? In fact, we can extract a huge amount of information from this kind of observations. Indeed, lensing is a unique tool for tracing the total mass distribution of the lens and at the same time it magnifies distant sources, allowing to see objects which otherwise would be very faint. Finally, since lensing is a geometrical effect, for lensing systems at cosmological distances some important information about the geometry of the universe can be derived.
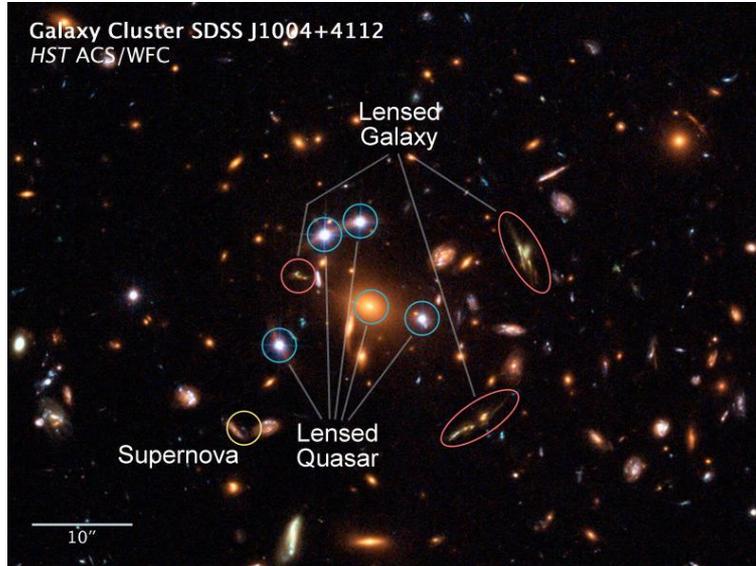
Figure 5.2: The galaxy cluster SDSS J1004+4112 has 5 images of a QSO close to its centre. A the same time, several other background galaxies are strongly lensed by this cluster.

It is then crucial to use in the proper way the observational constraints given a lens system.

Galaxies and galaxy clusters can strongly lens point or extended sources. The angular separation between multiple images if of the order of the Einstein ring,

$$
\begin{aligned}
\theta_E &= 0.9'' \left(\frac{M}{10^{11} M_\odot} \right)^{1/2} \left(\frac{D}{\text{Gpc}} \right)^{1/2} \\
&= 1.5' \left(\frac{M}{10^{15} M_\odot} \right)^{1/2} \left(\frac{D}{\text{Gpc}} \right)^{1/2},
\end{aligned}
\tag{5.8}
$$

which means that the multiple images can be well resolved.

Among the things that can be measured for a lens are

(1) the relative positions of the components (lens and images; astrometric constraints);

(2) the relative fluxes of the images;

(3) the time delays between the images;

(4) several properties of the lens (dynamical properties, light distribution in different bands, etc.);

(5) the microlensing of the images.

The astrometric constraints are the most important. We can usually measure the relative positions of the lensed components very accurately (5 mas or better). Substructures and other perturbers set a lower limit of order 1-5 mas with which it is safe to impose astrometric constraints. A given lens-source model provides a configuration of the

images which can be compared with the observations. If the images are extended, additional constraints derive from the fact that we can measure the relative transformation between one image and the other.

Flux ratios are relatively easy to measure, but not very useful because of the systematic uncertainties (see later in the text). As we will see, flux ratios are predictable. However, if taken at a sigle epoch they are affected by time variablity in the source (which appears with some time delay in the different images), microlensing (in galaxies) or lensing by subhalos and substructures (in galaxies and clusters), absorption by ISM (galaxies) and IGM (clusters). In fact, most applications of flux ratios have focused on probing these perturbers rather than on studying the mass distribution of the lenses.

Measurements of time delays are possible mainly for sources which are lensed by galaxies. Indeed, the angular separations between multiple images in galaxy clusters are typically too large and the time delays too long (except for close pairs of multiple images, like in SDSS J1004+4112, see astro-ph/0607513). For example, from Eqs. 2.83 and 3.69, the time delay between the two images produced by a SIS is

$$\Delta t_{SIS} = \frac{1}{2}\frac{D_\mathsf{L}D_\mathsf{S}}{cD_\mathsf{LS}}(1 + z_\mathsf{L})(\theta_A^2 - \theta_B^2) \ . \tag{5.9}$$

Such delay amounts to some months in the case of galaxies ($\theta_E \sim 1''$) and to some decades in the case of galaxy clusters. To date, time delays have primarily been used to estimate the Hubble constant rather than the surface density, but if we assume to know H_0 or consider only time delays ratios, then time delays can be used to constrain the mass distribution.

Any independent measurement of the mass of a component will also help to constrain the structure of the lens. For galaxies, this primarily means making stellar dynamical measurements of the lens galaxy and comparing the dynamical mass estimates to those from the lens geometry. For clusters, the dynamics of cluster galaxies, the X-ray emission from the hot intra-cluster gas, or weak lensing (see later) can be used to estimate the cluster mass. Unfortunately, these independent mass estimates are frequently in disagreement with those derived from strong lensing.

Using microlensing variability to constrain the mass distribution around where images appear is actually more theory than practice due to the lack of microlensing light curves for almost all lenses.

5.2.4 The importance of redshifts

There is an important piece of information about the lens and sources which is essential in order to extract any physical information from a lensing observation: redshifts. In order to transform angular quantities into masses and lengths, we need to know the redshift of both the lenses and the sources. Redshifts are not easy to obtain, because they require a spectroscopic follow-up. Especially for the source redshifts, long exposures with the largest telescopes are typically required. In many cases, spectroscopic redshifts are out of reach, and one must rely on photometric redshifts. In the case of lensing systems, these are not easy to measure, because of the light contamination by the foreground deflector. In galaxy clusters, this complication is particularly strong because the contaminants are many (cluster galaxies and intra-cluster light).

5.2.5 Lens optimization

The optimization procedure consists of finding the most probable set of parameters $\mathbf{p}$ which allow the model to reproduce the observables. In other words, it is a fitting

procedure, which is implemented defining a likelihood $\mathcal{L}$ for the observed data D and parameters $\vec{p}$,

$$\mathcal{L} = \mathrm{Pr}(D|\mathbf{p}) = \prod_{i=1}^{N} \frac{1}{\prod_{j=1}^{n_i} \sigma_{ij}\sqrt{2\pi}} \exp{-\frac{\chi_i^2}{2}} \; , \qquad (5.10)$$

where N is the number of systems of multiple images, n_i is the number of multiple images for the system i. The error on the constrain j belonging to the image system i is given by σ_{ij}. The optimization procedure must thus lead to the minimization of a χ^2 fit statistic.

For the astrometric constraints, the contribution from the multiple image system i to the overall χ^2 may be written as

$$\chi_i^2 = \sum_{j=1}^{n_i} \frac{[\vec{\theta}_{obs}^{\,j} - \vec{\theta}_{\mathbf{p}}^{\,j}]^2}{\sigma_{ij}^2} \; ; \; , \qquad (5.11)$$

where $\vec{\theta}_{obs}$ and $\vec{\theta}_{\mathbf{p}}$ are the observed and predicted positions of the images, and σ_{ij} are the positional errors.

The predicted positions $\vec{\theta}_{\mathbf{p}}$ must be derived from the lens equation. This involves two steps. In the first, we use the observed image positions to guess the position of their source:

$$\vec{\beta}_{\mathbf{p}}^{\,j} = \vec{\theta}_{obs}^{\,j} - \vec{\alpha}(\vec{\theta}_{obs}^{\,j}, \mathbf{p}) \; . \qquad (5.12)$$

In the second, we recast the source onto a set of multiple images on the image plane by solving the lens equation:

$$\vec{\beta}_{\mathbf{p}}^{\,j} = \vec{\theta}_{\mathbf{p}}^{\,j} - \vec{\alpha}(\vec{\theta}_{\mathbf{p}}^{\,j}, \mathbf{p}) \; . \qquad (5.13)$$

Since each image $\vec{\theta}_{obs}^{\,j}$ gives rise to a source $\vec{\beta}^{\,j}$, by remapping the source back onto the image plane, we ideally obtain a set n_i predicted multiple images for each observed $\vec{\theta}_{obs}^{\,j}$, which we can use for the optimization. In fact, depending on how far is the model from the truth, different image multiplicities may be predicted, so that the the χ^2 computation will fail and the corresponding model will be rejected.

Unfortunately, this procedure is very time consuming, because the lens equation must be solved numerically for complex mass distributions. One alternative approach is the optimization in the source plane. We can write the χ^2 as

$$\chi_{S,i}^2 = \sum_{j=1}^{n_i} \frac{[\beta_{\mathbf{p}}^j - \langle \beta_{\mathbf{p}}^j \rangle]}{\mu_j^{-2}\sigma_{ij}^2} \; , \qquad (5.14)$$

where $\langle \beta_{\mathbf{p}}^j \rangle$ is the barycenter of the positions of all the n_i source positions. Following this approach, the lens equation has not to be solved, so the computation is much faster. In the case of point sources, this works just fine, but it might lead to biases in the case of extended sources. Imagine that multiple conjugate features are recognized in the multiple images of an extended source. All these multiple knots can potentially be used as lensing constraints. The minimization of χ_S^2 actually leads to minimize the distance between the sources of the knots in the source plane, unless some regularization is assumed to prevent this. In such case, the optimization will tend to converge towards a solution which maximizes the magnification. In addition, the μ_j in Eq. 5.14 must be estimated from the model itself.

State of the art parametric modeling is done in the context of a Bayesian framework (Jullo et al., 2007), which is particularly useful when the data by themselves do not put

strong constraints on the model, which is quite common in the case of strong lensing modeling. The Bayes theorem states

$$\Pr(\mathbf{p}|D, M) = \frac{\Pr(D|\mathbf{p}, M)\Pr(\mathbf{p}|M)}{\Pr(D|M)} \ , \tag{5.15}$$

which reads: the likelihood of the parameters $\mathbf{p}$ given the model M and the data D (aka, the posterior PDF) is given by the likelihood of the data given the parameters and the model, times the likelihood of the parameters given the model (aka the prior PDF), divided by the likelihood of the data given the model (aka the evidence). Thus, the goal is to find the the set of parameters which maximizes the posterior PDF and which is consistent with prior PDF. This can be achieved by means of Monte Carlo Markov Chains to sample the parameter space until convergence.

5.2.6 Degeneracies

The mass modelization of the lens is never an easy task. In particular, the modeler has to deal with huge degeneracies in the parameter space which make several models able to fit the observed data. We cannot afford to discuss all possible model degeneracies, which also depend on the chosen parameterization. However, it is useful to consider some example.

Depending on the shape of its density profile, a lens can be more or lens sensitive to **external perturbations**. For example, an external shear $\vec{\gamma}_e$, which corresponds to a lensing potential

$$\Psi_\gamma(\vec{x}) = \frac{\gamma_e}{2}(x_1^2 - x_2^2) \ , \tag{5.16}$$

perturbs much more easily an NFW than a Singular-Isothermal sphere. The reason is that the SIS density profile is significantly steeper than the NFW in the central part. The situation is illustrated in Fig. 5.3. In the left panels, the critical lines and caustics of two axially symmetric lenses with the same mass ($M = 10^{14}\, M_\odot$) are shown. The lens and the source redshifts are assumed to be 0.3 and 1.0, respectively. An external shear, whose amplitude is $\gamma_e = 0.1$, is applied to both the lenses. The corresponding critical lines and caustics are shown in the two right panels. Clearly, the deformation of the critical lines (and caustics) is more significant for the NFW than for the SIS profile;

The **impact of the substructures** is therefore much larger in halos whose density profiles are shallower. The same two lenses which were earlier embedded into an external shear are now shown in Fig. 5.4 after having been populated with 30 axially symmetric subhalos. Each subhalo has mass between 10^{10} and $10^{11}\, M_\odot$ and has been modelled as a SIS. One could think that they represent galaxies in the halo of a galaxy cluster. The same subhalos have been used to populate both the lenses. Two important features can be noticed. First, for both the NFW and SIS profiles, the critical lines of the main clump are expanded. The relative expansion of the critical line of the NFW lens is significantly larger than that of the SIS. Substructures typically enhance the strong lensing hability of the lens, because they provide additional convergence and shear. Second, the critical lines around the individual subhalos have different sizes, depending on density profile of the halos within which they are embedded. In other words: substructures are much stronger lenses when the structure they belong to has a shallow density profile;

5.2.7 Free-form modelization

Another approach is the non-parametric one. The basic idea behind non-parametric mass models is that the effective lens potential and the deflection equations are linear