ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/225497437

# On modeling galaxy-scale strong lens systems

**Article** *in* General Relativity and Gravitation · September 2010

DOI: 10.1007/s10714-010-1041-1

CITATIONS
13

READS
42

**1 author:**

Charles R. Keeton
Rutgers, The State University of New Jersey
**180** PUBLICATIONS **5,472** CITATIONS

SEE PROFILE

All content following this page was uploaded by Charles R. Keeton on 23 December 2014.

The user has requested enhancement of the downloaded file.

Gen Relativ Gravit (2010) 42:2151–2176
DOI 10.1007/s10714-010-1041-1

REVIEW ARTICLE

# On modeling galaxy-scale strong lens systems

**Charles R. Keeton**

Received: 22 December 2009 / Accepted: 15 June 2010 / Published online: 30 June 2010
© Springer Science+Business Media, LLC 2010

**Abstract** I review methods for modeling gravitational lens systems comprising multiple images of a background source surrounding a foreground galaxy. In a Bayesian framework, the likelihood is driven by the nature of the data, which in turn depends on whether the source is point-like or extended. The prior encodes astrophysical expectations about lens galaxy mass distributions, either through a careful choice of model families, or through an explicit Bayesian prior applied to under-constrained free-form models. We can think about different lens modeling methods in terms of their choices of likelihoods and priors.

**Keywords** Strong gravitational lensing · Lens modeling · Galaxy structure · Statistical methods

## 1 Introduction

The gravitational deflection of light is a simple physical phenomenon with a rich phenomenology. The fact that many different observable effects arise from a few key physical principles gives gravitational lensing broad reach in astrophysics, cosmology, relativity, and even mathematics. While other contributions in this volume address different aspects of the phenomenology, I focus on strong lensing systems in which light bending by a distant, massive galaxy creates multiple, resolvable images of a more distant source.

Strong lensing supports a vast array of astrophysical applications, which are the subject of the comprehensive review by Kochanek [1] as well as other articles [2–5].

C. R. Keeton (✉)
Department of Physics and Astronomy, Rutgers, The State University of New Jersey,
136 Frelinghuysen Road, Piscataway, NJ 08854, USA
e-mail: keeton@physics.rutgers.edu

Rather than compete with those good expositions, I wish to highlight a core component of strong lensing studies: lens modeling. Lensing's boon—its ability to probe dark matter in distant galaxies—is also its bane, in the sense that we cannot see everything we need to understand in order to interpret lens data. We cannot avoid making some assumptions about how the mass in lens galaxies is distributed. To a large extent the assorted modeling methods that have emerged in recent years differ as to what assumptions are made and how they are used. That is what I aim to discuss.

As we will see, there can be assumptions about both the lens galaxy and the background source. With the source, the choice of point-like or extended is governed by the data, but if it is extended we need to decide what kind of symmetries and/or smoothness criteria to impose. With the lens, the fundamental (and contentious) choice is how to apply independent astrophysical knowledge. One approach is to define a modest space of models based on a careful analysis of what we think we know about the properties of galaxies that are important for lensing. A second approach is to give the lens models much more freedom and flexibility, and to supplement the lens data with explicit constraints on the form and/or smoothness of the lens model in order to obtain a reasonable set of acceptable models.

My goal is to review various lens modeling methods and see if we can discuss them in a common framework provided by Bayesian statistics. In Bayesian language, we draw conclusions from the posterior probability distribution for a model, which can be decomposed into two factors: the likelihood of the data given the model, and priors on the model. The form of the likelihood is guided by the nature of the data (see Sect. 4). The form of the priors, by contrast, depends on the choice of models. The two approaches mentioned above boil down to using astrophysical priors to narrow the model space before we encounter any lens data, or keeping the model space large and using priors in conjunction with lens data to constrain the posterior distribution. The distinction is often distilled into "parametric" versus "non-parametric" models, although I think it is more important to emphasize whether the models are over- or under-constrained in terms of the lens data (see Sect. 5).

Let me be clear at the outset about what I will and will not do. I will focus on strong lensing by galaxies, not galaxy clusters. Cluster lensing has its own distinctive features that I do not have space to address here (e.g., [6,7]). I will consider "macrolensing" by galaxies and their dark matter halos and "millilensing" by dark matter substructure, but will not include "microlensing" by stars in lens galaxies because that typically involves very different methods (see the contribution by Wambsganss in this volume). I will try to give enough technical detail about each modeling method to clarify how it works and make this presentation coherent, but leave the nuts and bolts of implementation to papers in the literature. I will not be able to cite all examples of different modeling methods; instead, I will try to cite the methods themselves and typical applications.

## 2 Basic lens theory

There are many good presentations of lens theory in different contexts (including the books by Schneider et al. [8] and Petters et al. [9], and other contributions in this

volume), but for completeness let me review the points that are important for strong lens modeling. I focus on the thin lens approximation with a single lens plane. I work with angular variables; the corresponding physical lengths in the lens or source plane may be obtained by multiplying by the appropriate angular diameter distance.

Consider a light ray that is emitted from angular position $\boldsymbol{u}$ in the source plane and passes through angular position $\boldsymbol{x}$ in the lens plane. Compared to an unlensed ray, the lensed ray has a longer travel time because it has a longer geometric length and it passes through a gravitational potential well. The excess light travel time is

$$t(\boldsymbol{x}) = t_0 \left[ \frac{1}{2} |\boldsymbol{x} - \boldsymbol{u}|^2 - \phi(\boldsymbol{x}) \right], \qquad t_0 = \frac{1+z_l}{c} \frac{D_{ol} D_{os}}{D_{ls}}, \tag{1}$$

where $z_l$ is the redshift of the lens, and $D_{ol}$, $D_{os}$, and $D_{ls}$ are angular diameter distances[1] from the observer to the lens, from the observer to the source, and from the lens to the source, respectively. We refer to the distance combination $D_{ol} D_{os} / D_{ls}$ as the "time-delay distance." The two-dimensional lens potential $\phi$ is given by the Poisson equation

$$\nabla^2 \phi = 2\kappa, \qquad \kappa = \frac{\Sigma}{\Sigma_{\rm crit}}, \qquad \Sigma_{\rm crit} = \frac{c^2}{4\pi G} \frac{D_{ol} D_{os}}{D_{ls}}. \tag{2}$$

The "convergence" $\kappa$ is the projected surface mass density of the lens, $\Sigma$, scaled by the critical density for lensing, $\Sigma_{\rm crit}$. Note that this expression for $\Sigma_{\rm crit}$ is in angular units; it yields the critical density in, say, $M_\odot\,{\rm arcsec}^{-2}$. The expression for $\Sigma_{\rm crit}$ in physical units (e.g., $M_\odot\,{\rm pc}^{-2}$) has $D_{ol}$ in the denominator rather than the numerator.

By Fermat's principle, images form at stationary points of the time delay surface, and the condition $\nabla t(\boldsymbol{x}) = 0$ yields the lens equation:

$$\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{\alpha}(\boldsymbol{x}), \qquad \boldsymbol{\alpha}(\boldsymbol{x}) = \nabla \phi(\boldsymbol{x}). \tag{3}$$

Strong lensing corresponds to a situation in which there are multiple solutions $\boldsymbol{x}_i$ of the lens equation for a given source position $\boldsymbol{u}$. The local distortion of an image is described by the magnification tensor,

$$\boldsymbol{\mu} \equiv \left( \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}} \right)^{-1} = \begin{bmatrix} 1 - \frac{\partial^2 \phi}{\partial x^2} & -\frac{\partial^2 \phi}{\partial x \partial y} \\ -\frac{\partial^2 \phi}{\partial x \partial y} & 1 - \frac{\partial^2 \phi}{\partial y^2} \end{bmatrix}^{-1}, \tag{4}$$

Since lensing conserves surface brightness, a change in flux can arise only from a change in the area of the image relative to the area of the source. For an unresolved source, the image and source plane areas are related by the scalar magnification $\mu = \det(\boldsymbol{\mu})$. In strong lensing there are critical curves in the lens plane along which the magnification tensor is singular ($\mu^{-1} = 0$), which map to caustics in the source plane. The caustics bound regions in the source plane that yield different numbers of images.

[1] The $1 + z_l$ factor in $t_0$ disappears if we use proper motion distances instead of angular diameter distances.

A source near a caustic produces images near a critical curve that are highly amplified and distorted. For a typical lens there are two critical curves, labeled tangential and radial based on the principal direction of the distortions.

We can write the potential and its derivatives as 2-d integrals over the surface mass density:

$$\phi(\boldsymbol{x}) = \frac{1}{\pi} \int \ln|\boldsymbol{x} - \boldsymbol{y}| \; \kappa(\boldsymbol{y}) \, d\boldsymbol{y}, \qquad \boldsymbol{\alpha}(\boldsymbol{x}) = \frac{1}{\pi} \int \frac{\boldsymbol{x} - \boldsymbol{y}}{|\boldsymbol{x} - \boldsymbol{y}|^2} \, \kappa(\boldsymbol{y}) \, d\boldsymbol{y}, \tag{5}$$

and there are similar expressions for the second derivatives of $\phi$. For ellipsoidal mass distributions, the symmetry allows the 2-d integrals to be reduced to 1-d integrals, which can sometimes be computed analytically but otherwise may be evaluated numerically [10,11]. For general mass distributions, one way to compute the lens potential and its derivatives is with Fourier techniques (see [12] for more discussion). Regardless of the symmetry of the lens galaxy (or lack thereof), a useful quantity is the monopole deflection,

$$\alpha_0(r) = \frac{1}{\pi r} \int_0^{2\pi} \int_0^r \kappa\left(r', \theta\right) \, r' \, dr' \, d\theta = \frac{M(r)}{\pi \Sigma_{\rm crit} r}, \tag{6}$$

where $M(r)$ is the projected mass enclosed by radius $r$. The monopole deflection curve is a 2-d analog of the rotation curve and therefore provides a useful way to characterize the density profile of a lens galaxy. Furthermore, the monopole deflection provides a general way to define the Einstein radius:

$$\alpha_0(R_{\rm ein}) = R_{\rm ein}. \tag{7}$$

The Einstein radius is one of the most robust constraints from lens models.

## 3 Lens modeling framework

Lens modeling amounts to using some data $d$ to draw inferences about some model $M$ specified by some parameters $\theta$. (The data and parameters may be multi-dimensional.) In the framework of Bayesian statistics [13], quantitative inferences are characterized by the posterior probability distribution for the parameters given the data,

$$P(\theta \,|\, d) = \frac{\mathcal{L}(d \,|\, \theta) \; P(\theta)}{P(d)}, \tag{8}$$

where $\mathcal{L}(d \,|\, \theta)$ is the likelihood of the data given the parameters, while the prior probability distribution $P(\theta)$ encodes what we believe about the parameters before encountering the data. The denominator,

$$P(d) = \int \mathcal{L}(d \,|\, \theta) \; P(\theta) \, d\theta, \tag{9}$$

is called the Bayesian evidence and is valuable when comparing different models. Before we discuss specific forms of the likelihood (Sect. 4) and priors (Sect. 5) for strong lensing, it is useful to review some basic principles of Bayesian inference.

To identify the "best" values of the parameters, we look for the peak in the posterior probability distribution. If we elect not to impose any explicit priors (which still amounts to a particular choice, often called "flat priors"), then finding the peak of the posterior is equivalent to a maximum likelihood analysis. We can often write the likelihood in the form $\mathcal{L} \propto \exp\left(-\chi^2/2\right)$ for an appropriate goodness-of-fit statistic $\chi^2$, so the best-fit parameters are associated with the minimum of the $\chi^2$ surface.

Finding the peak is not the end of the story, though; we also need to characterize the width of the posterior distribution to understand the uncertainties in the recovered parameters. If the posterior is Gaussian (so the $\chi^2$ is parabolic), we can estimate the uncertainties from the range of models that are within some $\Delta\chi^2$ threshold of the best fit [14]. The question arises of what to do if we are mainly interested in some parameter $\theta$ but our model also has a nuisance parameter $\lambda$. One strategy is to *optimize* $\lambda$: at each $\theta$ we pick the value of $\lambda$ that minimizes $\chi^2(\lambda \mid \theta)$, and we keep adjusting $\lambda$ as we scan through $\theta$.

The proper statistical approach, however, is to *marginalize* any nuisance parameter(s) by integrating:

$$P(\theta \mid d) = \int P(\theta, \lambda \mid d)\, d\lambda. \tag{10}$$

In general this is *not* the same as optimizing $\lambda$. Optimization is a reasonable proxy for marginalization only if the posterior is sharply peaked (i.e., nearly a $\delta$-function) in the $\lambda$ direction. In lens modeling the posterior is often peaked in certain directions but broad in others, so we use a judicious mix of optimization and marginalization (see Sects. 4 and 5).

When it is difficult to study the posterior analytically, we can resort to sampling the distribution. Sampling is versatile, especially in high dimensions, because it lets us construct various marginalized distributions just by binning the sampled points in different ways. Two popular sampling methods are Monte Carlo Markov Chains [13] and nested sampling [15,16].

In many lens modeling applications it is convenient to explore the parameter space using nested loops: an inner loop varies the source parameters while the lens model is held fixed; and then an outer loop varies the lens model parameters. One reason for doing this is to limit the number of times we must compute a new lens model (which may be the most computationally expensive step). A second reason is that in many applications the source parameters are not the main focus, so we might as well build their marginalization into the analysis. A third reason is that with source parameters we can often substitute optimization for marginalization; but the same cannot always be said of the lens parameters. I discuss source and lens model parameters separately in the next two sections.

## 4 Handling the sources

In the "source loop" we need to (1) use a (fixed) lens model to predict the image properties, (2) compute a likelihood to obtain a quantitative comparison between the model and the data, and (3) marginalize/optimize the source parameters. In this phase the methodology is pretty well determined by the nature of the data and does not involve much user choice (with one notable exception; see Sect. 4.5.2). While the methods vary between point sources (Sects. 4.1 and 4.2), elliptical sources (Sect. 4.3), and general extended sources (Sects. 4.4 and 4.5), there are three common principles. First, we can usually assume the data obey Gaussian statistics, which allows us to write the likelihood in the form $\mathcal{L} \propto \exp\left(-\chi^2/2\right)$ where $\chi^2$ is a familiar goodness-of-fit statistic; I give $\chi^2$ expressions here.[2] Second, if we have several independent datasets we can combine them by multiplying likelihood, or adding $\chi^2$'s. Third (as noted above), we typically consider optimization to be an acceptable substitute for marginalization when dealing with source parameters.

### 4.1 Point sources

If the source is point-like (such as a quasar), the data comprise positions, fluxes, and/or time delays for a discrete set of images, and the model includes the corresponding properties of the source. It is conceptually straightforward to write down the $\chi^2$ goodness of fit and discuss source optimization. The technical challenge is solving the lens equation to predict the model images.

#### *4.1.1 Making predictions*

The lens equation is usually non-linear, often transcendental, and may not even have a closed form (if the deflection integral in Eq. 5 cannot be evaluated). Needless to say, analytic solutions are rarely available. Numerical solutions can be hard to obtain because there is no algorithm that is guaranteed to find all the roots of a two-dimensional equation [14]. Root finders require independent knowledge of the number of images and rough guesses for their positions, but no local property of the lens equation gives this information. There are mathematical theorems about the *maximum* number of images that can be produced by given lens configurations (see [9,17–22] and the contribution by Petters and Werner in this volume), but we need to know the *actual* number of images produced by a source at a given position. The global caustic structure can be informative but hard to analyze.

To develop a general lens solver, let us read the lens equation "backwards" as a mapping from an image position $\boldsymbol{x}$ to a unique source position $\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\alpha}(\boldsymbol{x})$. If we lay down some tiling of the image plane, the lens equation maps each tile (vertex by vertex) to produce a tiling of the source plane that covers each point at least once. The number of tiles that cover the source reveals the number of images, and the tiles

[2] I give short shrift to the proportionality constant in $\mathcal{L} \propto \exp(-\chi^2/2)$; it is irrelevant when characterizing the peak and width of the posterior, and even when using evidence ratios to compare models.

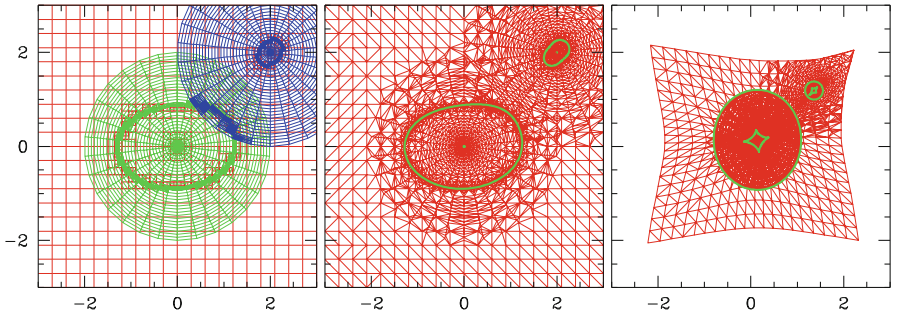


**Fig. 1** Illustration of the tiling algorithm for a main galaxy at the origin and a satellite nearby. The *left panel* shows a simple Cartesian grid (*red*) that serves as a foundation, plus polar grids (*green and blue*) centered on the two galaxies. All three use recursive subgridding near critical curves. The *middle panel* shows a Delaunay triangulation [26,27] of the grid points. The critical curves are now shown in *green*. The *right panel* shows the final grid (*red*) and caustics (*green*) in the source plane

themselves provide estimates of the image positions[3] that can be used to seed a numerical root finder. The tiling approach provides a fully general way to solve the lens equation [23–25], and it serves as the foundation for my GRAVLENS and LENSMODEL software.

There are two practical details to keep in mind. First, the tiles should be triangles so they remain convex regardless of how they are deformed by the lens mapping. Also, it is straightforward to determine whether a source point $\boldsymbol{u}$ is covered by a triangle with vertices $\boldsymbol{u}_i$ $(i = 1, 2, 3)$: given $\delta\boldsymbol{u}_i = \boldsymbol{u}_i - \boldsymbol{u}$, the source lies within the triangle if the three cross products $\delta\boldsymbol{u}_1 \times \delta\boldsymbol{u}_2$, $\delta\boldsymbol{u}_2 \times \delta\boldsymbol{u}_3$, and $\delta\boldsymbol{u}_3 \times \delta\boldsymbol{u}_1$ all have the same sign. Second, the tiling has finite resolution, but we can use an adaptive algorithm to obtain high resolution where appropriate. For example, we might combine a coarse Cartesian grid (for basic coverage) with polar grids centered on any or all galaxies in the model (to capture their features). To resolve the folding and distortion near critical curves, we can recursively subdivide any tile in which the scalar magnification $\mu$ changes sign (indicating that a critical curve traverses the tile). Once we have the desired grid points, we can combine them efficiently using Delaunay triangulation [26,27]. A sample tiling is shown in Fig. 1.

### *4.1.2 Handling data*

In this presentation I assume the lens is well resolved so there are no correlations between the measurement uncertainties for different images, but the framework can be extended to handle correlated errors. I also assume that either the fluxes have been corrected for time delays and possible extinction in the lens galaxy, or the errorbars have been broadened to account for those effects.

[3] If source plane tile $S_k$ covers the source position, then the corresponding image plane tile $I_k$ must contain an image. The only subtlety involves edges: the straight edges of $S_k$ may not map exactly to the assumed straight edges of $I_k$ (or vice versa). So the actual image position could lie slightly outside $I_k$, but it ought to be close.

The $\chi^2$ term for the image positions has the form [28]

$$\chi^2_{\rm pos} = \sum_i \left(\boldsymbol{x}_i^{\rm mod} - \boldsymbol{x}_i^{\rm obs}\right)^t \mathbf{S}_i^{-1} \left(\boldsymbol{x}_i^{\rm mod} - \boldsymbol{x}_i^{\rm obs}\right), \tag{11}$$

$$\approx \sum_i \left(\boldsymbol{u}^{\rm mod} - \boldsymbol{u}_i^{\rm obs}\right)^t \boldsymbol{\mu}_i^t \mathbf{S}_i^{-1} \boldsymbol{\mu}_i \left(\boldsymbol{u}^{\rm mod} - \boldsymbol{u}_i^{\rm obs}\right), \tag{12}$$

where the sum runs over the images and the astrometric uncertainties are described by the covariance matrix

$$\mathbf{S}_i = \mathbf{R}_i \begin{bmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{bmatrix} \mathbf{R}_i^t, \qquad \mathbf{R}_i = \begin{bmatrix} -\sin\theta_{\sigma i} & -\cos\theta_{\sigma i} \\ \cos\theta_{\sigma i} & -\sin\theta_{\sigma i} \end{bmatrix}, \tag{13}$$

where we allow an error ellipse with semi-major axis $\sigma_{1i}$, semi-minor axis $\sigma_{2i}$, and position angle $\theta_{\sigma i}$ (measured East of North). If there are multiple sets of images corresponding to different sources, then $\chi^2_{\rm pos}$ contains a set of terms for each multiply-imaged system. In Eq. (12), $\boldsymbol{u}_i^{\rm obs}$ is the projection of the observed position of image $i$ into the source plane,

$$\boldsymbol{u}_i^{\rm obs} = \boldsymbol{x}_i^{\rm obs} - \boldsymbol{\alpha}\left(\boldsymbol{x}_i^{\rm obs}\right), \tag{14}$$

and we use the fact that a small displacement $\delta\boldsymbol{x}_i = \boldsymbol{x}_i^{\rm mod} - \boldsymbol{x}_i^{\rm obs}$ in the image plane can be related to a small displacement $\delta\boldsymbol{u}_i = \boldsymbol{u}^{\rm mod} - \boldsymbol{u}_i^{\rm obs}$ in the source plane via a Taylor series expansion: $\delta\boldsymbol{x}_i \approx \boldsymbol{\mu}_i\,\delta\boldsymbol{u}_i + \cdots$ (with $\boldsymbol{\mu}_i$ evaluated at the observed image position). This approximation is valid only when the position residuals are small, so strictly speaking Eq. (12) may not be accurate when the residuals are large; but this is not a big concern, because the $\chi^2_{\rm pos}$ value will still indicate a bad fit. There are two reasons why Eq. (12) may be attractive. First, it avoids the need to solve the lens equation: we only project from the lens plane to the source plane and never need to go the other way. Second, with Eq. (12) we can find the optimal source position analytically:[4]

$$\boldsymbol{u}^{\rm mod} = \mathbf{A}^{-1}\,\boldsymbol{b}, \qquad \mathbf{A} = \sum_i \boldsymbol{\mu}_i^t\,\mathbf{S}_i^{-1}\,\boldsymbol{\mu}_i, \qquad \boldsymbol{b} = \sum_i \boldsymbol{\mu}_i^t\,\mathbf{S}_i^{-1}\,\boldsymbol{\mu}_i\,\boldsymbol{u}_i^{\rm obs}. \tag{15}$$

There is one concern when using Eq. (12): since we do not actually solve the lens equation, we cannot check whether the model predicts too many images. (Models that predict too few images are less worrisome, because $\boldsymbol{u}_i^{\rm obs}$ for the unpredicted image(s) would lie away from the others and lead to a large $\chi^2$ value.) The choice between Eqs. (11) and (12) depends, therefore, on the importance of verifying the number of model images versus obtaining faster runtimes. That choice may vary from one application to another.

[4] Here we assume the peak of the likelihood for $\boldsymbol{u}^{\rm mod}$ is dominated by the position constraints and not very sensitive to flux constraints.

The $\chi^2$ term for the image fluxes typically has the form

$$\chi^2_{\rm flux} = \sum_i \frac{\left(F_i^{\rm obs} - \mu_i F^{\rm src}\right)^2}{\sigma^2_{f,i}}. \tag{16}$$

If desired, parity information can be included by giving $F_i^{\rm obs}$ a sign that reflects the image parity (which can usually be inferred from the image configuration [29]), and letting the scalar magnification $\mu_i$ be signed as well. The optimal source flux can be found analytically,

$$F^{\rm src} = \frac{\sum_i F_i^{\rm obs}\,\mu_i/\sigma^2_{f,i}}{\sum_i \mu_i^2/\sigma^2_{f,i}}. \tag{17}$$

If it is preferable to express the photometric errors in magnitudes, this analysis may be generalized by writing the predicted magnitude as $m_i^{\rm mod} = m^{\rm src} - 2.5\log|\mu_i|$ and fitting for $m^{\rm src}$. Note that the photometric units are arbitrary; they can be absolute fluxes or magnitudes, or they can be relative fluxes or differential magnitudes referenced to one of the images.

Before discussing the $\chi^2$ for time delays, let us rewrite Eq. (1) to express the predicted time delay for image $i$ as

$$t_i^{\rm mod} = t_0\tau_i^{\rm mod} + T_0, \qquad \tau_i^{\rm mod} = \frac{1}{2}\left|\boldsymbol{x}_i^{\rm mod} - \boldsymbol{u}^{\rm mod}\right|^2 - \phi\left(\boldsymbol{x}_i^{\rm mod}\right), \tag{18}$$

where $t_0$ is again the time-delay distance in time units, and $T_0$ is a time zeropoint. Historically people have used the fact that $t_0 \propto H_0^{-1}$ to constrain the Hubble constant, but formulating the problem in terms of time-delay distances may enable additional cosmological applications [30–32]. The zeropoint $T_0$ does not affect differential time delays, of course, but I include it to make the framework general. Using Eq. (18) we can write a $\chi^2$ term of the form

$$\chi^2_{\rm tdel} = \sum_i \frac{\left(t_i^{\rm obs} - t_0\tau_i^{\rm mod} - T_0\right)^2}{\sigma^2_{t,i}}. \tag{19}$$

If we know the lens and source redshifts and have priors on the cosmological parameters (including $H_0$), we can turn them into a prior $t_{0,\rm prior} \pm \sigma_{t0}$ on the time scale and then impose them via a $\chi^2$ term

$$\chi^2_{t0} = \frac{(t_0 - t_{0,\rm prior})^2}{\sigma^2_{t0}}. \tag{20}$$

We can find the optimal values of $t_0$ and $T_0$ by solving the matrix equation

$$\begin{bmatrix} \sum_i \frac{(\tau_i^{\rm mod})^2}{\sigma_{t,i}^2} + \frac{1}{\sigma_{t0}^2} & \sum_i \frac{\tau_i^{\rm mod}}{\sigma_{t,i}^2} \\ \sum_i \frac{\tau_i^{\rm mod}}{\sigma_{t,i}^2} & \sum_i \frac{1}{\sigma_{t,i}^2} \end{bmatrix} \begin{bmatrix} t_0 \\ T_0 \end{bmatrix} = \begin{bmatrix} \sum_i \frac{\tau_i^{\rm mod} t_i^{\rm obs}}{\sigma_{t,i}^2} + \frac{t_{0,\rm prior}}{\sigma_{t0}^2} \\ \sum_i \frac{t_i^{\rm obs}}{\sigma_{t,i}^2} \end{bmatrix}. \tag{21}$$

## 4.2 Point sources: alternate approach

I briefly mention a rather different approach that has been used to find broad ranges of lens potentials (as opposed to specific models) that are consistent with observed lens configurations. The idea is to characterize a pair or triplet of images with just a few quantities that describe the configuration: the separation and opening angle [33], asymmetry and opening angle [34], or (for quad lenses) the separation between a pair of images and distance to the next nearest image [35,36] or combinations of image angles [37]. We say a model lens matches an observed one if the configuration quantities agree within some small tolerance. We set the likelihood to be uniform for models that match an observed lens, and zero otherwise. The utility of this simple approach is discussed in Sect. 5.5.

## 4.3 Elliptical sources

The issues are rather different when the source is extended. Consider a source with surface brightness distribution $S(\boldsymbol{u})$. Since lensing conserves surface brightness, we can immediately write down the form of the surface brightness distribution in the image plane:

$$I(\boldsymbol{x}) = S(\boldsymbol{u}(\boldsymbol{x})) = S(\boldsymbol{x} - \boldsymbol{\alpha}(\boldsymbol{x})). \tag{22}$$

Note that we do not need to solve the lens equation here; we just use the lens mapping to go from $\boldsymbol{x}$ to $\boldsymbol{u}(\boldsymbol{x})$. The challenge is not technical but conceptual: How do we identify the "information content" of the image $I(\boldsymbol{x})$? What features of the image help constrain the lens model?

### *4.3.1 Making predictions*

Blandford et al. [38] and Suyu and Blandford [39] point out that if you have a true surface brightness map (i.e., one not smeared by a PSF), and you see isophotes that cross, that tells you two things. First, the crossing point must lie on a critical curve. Second, the corresponding isophote in the source must be tangential to the caustic. These are certainly interesting, but they provide only local information and depend on having a clean surface brightness map, so they are mainly useful as qualitative tests of a lens model (see [39]).

Kochanek et al. [40] make things more quantitative by defining the "ring curve" to be the collection of points at which $I(\boldsymbol{x})$ has a local maximum in the radial direction.

To find the ring curve, pick some reference point $\boldsymbol{x}_0$ (e.g., the center of the lens galaxy) as the origin of polar coordinates $(r, \theta)$. For each azimuth $\theta$, find the point at which the radial gradient vanishes, $\hat{\boldsymbol{r}}^t \nabla I = 0$, where $\hat{\boldsymbol{r}} = \cos\theta\, \hat{\boldsymbol{x}} + \sin\theta\, \hat{\boldsymbol{y}}$ is the radial unit vector. We can understand the ring curve theoretically if the source has elliptical symmetry, which means the source surface brightness distribution can be written as $S(m(\boldsymbol{u}))$ where $m$ is the elliptical radius. Let the source be centered at $\boldsymbol{u}_0$ with axis ratio $q_s \leq 1$ and position angle $\theta_s$ (measured East of North), and construct the shape matrix

$$\mathbf{S}_s = \mathbf{R}_s \begin{bmatrix} 1 & 0 \\ 0 & q_s^2 \end{bmatrix} \mathbf{R}_s^t, \qquad \mathbf{R}_s = \begin{bmatrix} -\sin\theta_s & -\cos\theta_s \\ \cos\theta_s & -\sin\theta_s \end{bmatrix}. \tag{23}$$

In coordinates $(u', v')$ centered on the source and aligned with the major axis, the elliptical radius has $m^2 = (u')^2 + (v')^2/q_s^2$. In a general coordinate system,

$$m^2 = (\boldsymbol{u} - \boldsymbol{u}_0)^t\, \mathbf{S}_s^{-1}\, (\boldsymbol{u} - \boldsymbol{u}_0). \tag{24}$$

We assume $S(m)$ is a monotonic decreasing function but otherwise do not specify the brightness profile. With $I(\boldsymbol{x}) = S(m(\boldsymbol{x} - \boldsymbol{\alpha}))$, the equation for the ring curve becomes [40]

$$\hat{\boldsymbol{r}}^t \nabla I = 0 \quad \Rightarrow \quad (\boldsymbol{x} - \boldsymbol{\alpha} - \boldsymbol{u}_0)^t\, \mathbf{S}_s^{-1}\, \boldsymbol{\mu}^{-1}\, \hat{\boldsymbol{r}} = 0. \tag{25}$$

The surface brightness profile has dropped out: the only property of the source that affects the ring curve is its angular shape. The ring curve can now be found with simple 1-d numerical root finding along each spoke. If desired, we can then consider the brightness *along* the ring: local maxima correspond to images of the source center, while local minima correspond to positions at which the ring curve crosses the critical curve [38–40].

#### *4.3.2 Handling data*

Suppose we have an Einstein ring image whose ring curve is characterized by radii $r_i^{\rm obs} \pm \sigma_{r,i}$ at azimuthal angles $\theta_i$. We can then construct the $\chi^2$ term

$$\chi^2_{\rm ring} = \sum_i \frac{\left[r_i^{\rm obs} - r^{\rm mod}(\theta_i)\right]^2}{\sigma_{r,i}^2}. \tag{26}$$

This is a non-linear function of the source position, axis ratio, and position angle, so we must do an explicit search to optimize those parameters. If we measure the locations of the maxima and minima in the brightness along the ring, we can add additional terms imposing those constraints. Kochanek et al. [40] caution, however, that brightness minima may be difficult to measure and subject to systematic uncertainties such as seeing and dust extinction.

Several studies have found ring curves to be quite valuable for constraining not only the source shape but also the lens model [40–43]. The analysis is not hindered by moderate PSF smearing, but it does depend on the assumption of elliptical symmetry in the source, which may be reasonable in some circumstances but too restrictive in others.

If we have both an Einstein ring and point-like images, we can combine constraints by simply adding $\chi^2$ terms. There may be two challenges, though. On the technical side, it may be difficult to trace the ring curve underneath bright point images. On the conceptual side, it may not be obvious how many points along the ring to measure. In principle $N_{\rm ring}$ could be as large as the number of resolution elements along the ring, but that is not a precise number. The issue is important because increasing $N_{\rm ring}$ gives the ring constraints more weight relative to the point-image constraints.

### 4.4 Extended sources as smeared collections of point sources

In more general cases we cannot avoid making a detailed model of the source. One approach is to approximate an extended source as a collection of point sources such that smearing by the PSF (or beam in radio parlance) yields a smooth image. The task of modeling is to find the positions and fluxes for an appropriate set of point sources. This approach has been developed by Kochanek, Narayan, Ellithorpe, and Wucknitz in the LensClean algorithm for fitting lensed radio rings [44–46].

The set of point sources is built iteratively: at each step we add one more source whose position and flux are chosen to minimize the mean square residuals. Heuristically, we want to add a new component at the source position corresponding to the highest peak in the previous image plane residuals (see [44]). The details are more complicated in practice, because with radio observations it is preferable to work directly with the visibility data; see [45,46] for full discussions. There are two main points I want to emphasize here. First, in this approach it is not necessary to make any particular assumptions about the form of the source; that emerges naturally from the iterative approach. Second, this approach can be computationally intensive because at each step we need to use methods from Sect. 4.1.1 to find and account for all the images of the new source.

### 4.5 Pixelated sources

Another way to deal with general extended sources is to pixelate both the source and image. This leads to an attractive framework that is applicable to optical and infrared observations, or to radio maps that have already been CLEANed.

#### *4.5.1 Making predictions*

Let $\boldsymbol{d}$ and $\boldsymbol{s}$ be vectors containing the image and source plane pixel values, respectively.[5] The pixelization schemes need not be the same in the image and source planes.

[5] While we normally think of an observed image as an *array* of pixels, we can construct an equivalent *vector* by placing the rows (or columns) one after another.

The value $d_i$ at position $\boldsymbol{x}_i$ can be found by interpolating among source pixels to estimate the source surface brightness at $\boldsymbol{u}(\boldsymbol{x}_i)$. In other words, each $d_i$ can be written as a linear combination of $s_j$'s for some (usually modest) set of source pixels: $d_i = \sum_j (L_0)_{ij}\, s_j$. In matrix form this becomes $\boldsymbol{d} = \mathbf{L}_0 \boldsymbol{s}$ where the "lensing operator" $\mathbf{L}_0$ is a (sparse) matrix that encodes the lens mapping.

That analysis gives a pure surface brightness map, but it can be extended to include PSF smearing. Convolution is a linear operation, so we can discretize it with another linear operator $\mathbf{B}$. Then the final predicted image, incorporating both lensing and PSF effects, can be written as

$$\boldsymbol{d} = \mathbf{L}\,\boldsymbol{s}, \tag{27}$$

where the "blurred lensing operator" is $\mathbf{L} = \mathbf{B}\,\mathbf{L}_0$. This approach has been used by a variety of studies, sometimes with uniform source plane pixels and other times with source plane pixels that vary depending on the lensing magnification [47–56]. The upshot is that once you invest the time to construct the operator $\mathbf{L}$ for a given lens model (and PSF), you can quickly and easily construct images for arbitrary source models.

### *4.5.2 Handling data*

Suppose we have an extended image with data vector $\boldsymbol{d}^{\rm obs}$ whose uncertainties are described by the covariance matrix $\mathbf{S}_d$. Since we can write the predicted image as $\boldsymbol{d}^{\rm mod} = \mathbf{L}\,\boldsymbol{s}$, we can define the $\chi^2$ as

$$\chi^2_{\rm img} = \left(\mathbf{L}\,\boldsymbol{s} - \boldsymbol{d}^{\rm obs}\right)^t\, \mathbf{S}_d^{-1}\, \left(\mathbf{L}\,\boldsymbol{s} - \boldsymbol{d}^{\rm obs}\right). \tag{28}$$

In general it is necessary or at least desirable to supplement these direct constraints with some priors on the source model. It is popular to eschew direct priors in favor of regularization—to let the source remain free but penalize models that are too spiky or otherwise unrealistic. This effectively applies Occam's razor to trim away unnecessary complications and leave models that strike a balance between fit quality and complexity (see [54]).

There has been considerable discussion in the literature about how to regularize source models in lens modeling. One choice is to maximize the entropy $-\sum_j s_j \ln s_j$ of the source surface brightness distribution [48–50]. Combining this with Eq. (28) means we want to minimize

$$\chi^2 = \left(\mathbf{L}\,\boldsymbol{s} - \boldsymbol{d}^{\rm obs}\right)^t\, \mathbf{S}_d^{-1}\, \left(\mathbf{L}\,\boldsymbol{s} - \boldsymbol{d}^{\rm obs}\right) + \lambda_s \sum_j s_j \ln s_j, \tag{29}$$

where $\lambda_s$ controls the strength of the regularization such that a higher $\lambda_s$ gives more weight to the regularization whereas a lower $\lambda_s$ gives more weight to the quality of the fit. Entropy-based regularization automatically imposes the condition that source pixels should not be negative. This approach has met with success, although it can

be computationally expensive because it requires an explicit search of the (high-dimensional) $\boldsymbol{s}$ parameter space.

As an alternative, suppose we just try to avoid spikes by penalizing pixel values far from zero through the penalty function $\sum_j s_j^2 = \boldsymbol{s}^t \boldsymbol{s}$. Or we might penalize large gradients in the surface brightness distribution. Using finite differencing, we can estimate the gradient at each pixel using a linear combination of nearby pixels; so we put $\boldsymbol{v} = \mathbf{H}_s \boldsymbol{s}$ where the linear operator $\mathbf{H}_s$ encodes the finite differencing, and then use the penalty $\boldsymbol{v}^t \boldsymbol{v}$. Yet another possibility is to penalize large curvature terms; and the curvature can again be estimated via a linear operation on $\boldsymbol{s}$. These examples show that there are many different types of regularizations that can all be imposed through a quadratic penalty such that the total $\chi^2$ has the form [51–56]

$$\chi^2 = \left(\mathbf{L}\,\boldsymbol{s} - \boldsymbol{d}^{\rm obs}\right)^t \mathbf{S}_d^{-1} \left(\mathbf{L}\,\boldsymbol{s} - \boldsymbol{d}^{\rm obs}\right) + \lambda_s\, \boldsymbol{s}^t\, \mathbf{H}_s^t\, \mathbf{H}_s\, \boldsymbol{s}. \tag{30}$$

See [51,54] for explicit examples of $\mathbf{H}_s$ for amplitude-, gradient-, and curvature-based regularization schemes.

A weakness of quadratic regularization is that it does not forbid negative source pixels. A benefit is found in the fact that Eq. (30) is quadratic in $\boldsymbol{s}$, so we can find the optimal source model analytically by solving

$$(\mathbf{L}^t\, \mathbf{S}_d^{-1}\, \mathbf{L} + \lambda_s\, \mathbf{H}_s^t\, \mathbf{H}_s)\, \boldsymbol{s} = \mathbf{L}^t\, \mathbf{S}_d^{-1}\, \boldsymbol{d}^{\rm obs}. \tag{31}$$

While this expression may look a bit complicated, it is just a linear equation for $\boldsymbol{s}$ that can be solved with standard matrix routines [14]. To some people, the benefit that quadratic regularization significantly reduces the computational effort needed to reconstruct the source outweighs the cost of not excluding negative pixels.

All told, there are three choices to be made with regularized, pixelated sources: the source plane pixelization,[6] the form of regularization, and the strength of regularization. First consider the strength. In the Bayesian framework we can treat $\lambda_s$ as a nuisance parameter and marginalize it (perhaps with a flat prior in $\log \lambda_s$, where we use the logarithm because we do not know the natural scale). In practice, the probability is expected to have a sharp peak in the $\lambda_s$ direction, so we can just set $\lambda_s$ to the value at the peak. The peak can be found either by explicitly scanning in $\lambda_s$ or by solving an appropriate (non-linear) equation [54]. As for the form of the regularization, Suyu et al. [54] discuss how to use the Bayesian evidence to objectively rank different regularization schemes.

## 5 Handling the lenses

In the "lens loop" we need to vary the lens parameters to explore the posterior probability distribution. In the Bayesian framework, the posterior involves not just the likelihood (see Sect. 4) but the priors as well. Priors play a significant role in lens modeling, whether explicitly or implicitly. In traditional parametric lens modeling

[6] N.B. The image plane pixelization is usually dictated by the data.

(Sect. 5.1) the priors may not be apparent, but they are present nonetheless: they enter through the choice of density distributions used to fit the data. Advocates say this approach provides a good way to build astrophysical knowledge into lens modeling to ensure that the final models are plausible. The analysis of millilensing by dark matter substructure (Sect. 5.2) shows that a Bayesian approach to parametric modeling can be made quite rich. Some have questioned, though, whether modern parametric models are still too simple compared with real lens mass distributions, and have developed various methods that provide much more freedom (Sects. 5.3 and 5.4). Bayesian priors play a much more explicit role in such "free-form" and "hybrid" models. Finally, for certain applications it is valuable to do as little modeling as possible and try to find the full range of properties that can be produced by "realistic" mass distributions (Sect. 5.5). Such cases obviously rely heavily on priors to define what we mean by "realistic."

### 5.1 Parametric models

The oldest approach to lens modeling is to assume the density distributions of lens galaxies can be approximated by functions with a modest number of free parameters. Astrophysical knowledge is built into the careful selection of functions and parameters, in two ways. First, we can consider density distributions derived from other observational and theoretical studies of galaxies: the list includes isothermal, power law, de Vaucouleurs, Hernquist, and NFW models, to name a few (see [11] for a catalog). Second, we can study how various properties of lens galaxy mass distributions affect lensing observables and try to include all the properties that are important for a given dataset. This can be empirical in the sense of trying a number of different functional forms to identify both universal features and systematic uncertainties (e.g., [28,57]), or more formal (e.g., [40,58,59]).

There are a number of general principles that have emerged from parametric lens modeling:

- The Einstein radius, and the mass within it, are robust measurements with model uncertainties at the percent level [28,60].
- It can be difficult to constrain the radial density profile from strong lensing alone, especially in 4-image lenses [57–59]. Nevertheless, combining strong lensing with other astrophysical evidence indicates that the total density profiles of lens galaxies are close to isothermal (i.e., $\Sigma \propto r^{-1}$ in 2-d, or $\rho \propto r^{-2}$ in 3-d), although they might not be perfectly isothermal at all radii and there may be some intrinsic scatter [12,61–63].
- We cannot assume lens galaxies are spherical or isolated. Image position data, especially in 4-image lenses and rings, are good enough to detect and even distinguish quadrupole moments due to ellipticity in the lens galaxy and tidal shear from the lens galaxy's environment [64].
- Smooth lens models often fail to reproduce the observed flux ratios in 4-image lenses. The failures can usually be attributed to elements missing from the smooth models, namely microlensing by stars and/or millilensing by dark matter substructure [65–68].

Let me say a few more words about lens environments in connection with the third point. Many lens galaxies lie in groups or clusters of galaxies or have such massive structures projected along the line of sight [69,70]. Those structures can affect the light bending in ways we cannot ignore. If the perturbing mass lies outside the Einstein radius, we can approximate its contributions to the lens potential using a Taylor series expansion of the form[7]

$$\phi_{\rm env} = \frac{\kappa_c}{2} r^2 + \frac{\gamma}{2} r^2 \cos 2(\theta - \theta_\gamma) + \frac{\sigma}{4} r^3 \cos(\theta - \theta_\sigma) + \frac{\delta}{6} r^3 \cos 3(\theta - \theta_\delta) + \cdots \tag{32}$$

Here $\kappa_c$ represents the focusing due to a uniform mass sheet. This term can be absorbed by rescaling the lens potential due to the galaxy, such that the only observable effect is on time delays; this creates the (in)famous "mass sheet degeneracy" in lensing [71,72]. Also, $\gamma$ and $\theta_\gamma$ quantify the tidal shear created by an asymmetric distribution of mass in the lens environment; we can equivalently define $\gamma_c = \gamma \cos 2\theta_\gamma$ and $\gamma_s = \gamma \sin 2\theta_\gamma$. The external shear is often non-negligible in real lens systems, and detectable with modern data, so it represents a sort of minimal environment model. If the lens lies in a group or cluster of galaxies, such that there are perturbers within a few tens of Einstein radii, the higher-order terms may be non-negligible as well [73,74].

In the context of this review, I want to emphasize that parametric lens models can be used with all of the different source modeling methods discussed in Sect. 4. Indeed, parametric lens models have been used with point-like sources [28,57,61,63], with Einstein rings modeled using elliptical sources [40,41], with extended sources in LensClean [44–46], and with pixelated sources [47–50,75]. The principles discussed above have been drawn from the full range of parametric lens modeling studies.

One final point about parametric lens modeling is that we do not always think about it in a Bayesian sense because we do not necessarily have to deal with explicit priors. Nevertheless, it is straightforward and even attractive to place parametric modeling in a Bayesian framework. We might want to do that in order to impose additional astrophysical constraints (for example, on the shape and mass-to-light ratio of the stellar distribution if we are modeling the stars and dark matter separately). Or we might want a formal way to deal with nuisance parameters, as discussed next.

### 5.2 Parametric substructure models

The analysis of dark matter substructure by Dalal and Kochanek [68] provides a rich example of parametric modeling in a Bayesian framework. I describe a somewhat more general framework and then discuss what Dalal and Kochanek actually did. If we treat the lens as some smooth mass distribution plus a collection of clumps, we have the following parameter list:

– $\boldsymbol{q}_{\rm gal}$: parameters for the smooth component

[7] We neglect the zeroth-order term, which represents the zeropoint of the potential, and the first-order terms, which represent an unobservable translation of the source plane.

- $\boldsymbol{q}_{\rm sub}$: parameters for the *population* of clumps, describing the mean density of substructure as a function of position, and the clump mass function
- $\{\boldsymbol{x}_\alpha, m_\alpha\}$: all the individual clump positions and masses

Because the clump positions and masses are unknown, the number of parameters (far) exceeds the number of constraints. We do not actually care about the individual clump properties, though, so we can treat them as nuisance parameters and marginalize them. Formally, the posterior for the interesting parameters can be written as an integral over all clump positions and masses:

$$P(\boldsymbol{q}_{\rm gal}, \boldsymbol{q}_{\rm sub} \,|\, d) \propto \int \mathcal{L}(d \,|\, \boldsymbol{q}_{\rm gal}, \boldsymbol{q}_{\rm sub}, \{\boldsymbol{x}_\alpha, m_\alpha\})\ P(\{\boldsymbol{x}_\alpha, m_\alpha\} \,|\, \boldsymbol{q}_{\rm sub}) \times P(\boldsymbol{q}_{\rm sub} \,|\, \boldsymbol{q}_{\rm gal})\ P(\boldsymbol{q}_{\rm gal})\ d\{m_\alpha, \boldsymbol{x}_\alpha\}. \tag{33}$$

The first term is the likelihood for a given set of clumps. The second term is the probability distribution for the clump positions and masses given the population parameters. The third term represents priors on the substructure population parameters, which may depend on the galaxy parameters if we think the substructure should be related to the smooth component. Finally, the fourth term represents priors on the smooth component.

We have little chance of evaluating the high-dimensional integral directly, of course. Instead, we can compute the integral with Monte Carlo techniques: we generate many random realizations of the clump positions and masses drawn from the probability distribution $P(\{\boldsymbol{x}_\alpha, m_\alpha\} \,|\, \boldsymbol{q}_{\rm sub})$, and then replace the integral with a sum over those realizations. If $\{\boldsymbol{x}_\alpha, m_\alpha\}_\nu$ denotes the full set of clump positions and masses for the random realization with index $\nu$, then Eq. (33) becomes

$$P(\boldsymbol{q}_{\rm gal}, \boldsymbol{q}_{\rm sub} \,|\, d) \propto \left[\sum_{\nu=1}^{N_{\rm pop}} \mathcal{L}(d \,|\, \boldsymbol{q}_{\rm gal}, \boldsymbol{q}_{\rm sub}, \{\boldsymbol{x}_\alpha, m_\alpha\}_\nu)\right] P(\boldsymbol{q}_{\rm sub} \,|\, \boldsymbol{q}_{\rm gal})\ P(\boldsymbol{q}_{\rm gal}), \tag{34}$$

where the sum runs over the $N_{\rm pop}$ realizations of the clump population. (Note that $P(\{\boldsymbol{x}_\alpha, m_\alpha\} \,|\, \boldsymbol{q}_{\rm sub})$ does not appear explicitly in Eq. 34 because it is used to generate the substructure realizations.)

Dalal and Kochanek [68] treated the smooth component as an isothermal ellipsoid with external shear (with flat priors), and they assumed the clump population was spatially uniform with mean surface mass density $\Sigma_{\rm sub}$ such that all clumps had the same mass $m$. They were mainly interested in $\Sigma_{\rm sub}$, and to a lesser extent in $m$ (it was less well constrained), so in some sense even the smooth model parameters were a nuisance. While they did not fully marginalize the smooth model parameters, they did reoptimize them for each substructure realization. (They showed that the reoptimization can be linearized under the assumption that substructure perturbations to the potential and deflection are small.)

### 5.3 Free-form models

An alternative to selecting parametric models is to expand the lens potential using some appropriate set of basis functions,

$$\phi(\boldsymbol{x}) = \sum_{\nu} a_\nu \, f_\nu(\boldsymbol{x}), \tag{35}$$

and then fit for the coefficients $a_\nu$. While such models have sometimes been called "non-parametric," I consider that term misleading because there are in fact parameters (namely the expansion coefficients). I prefer the term "free-form" because it highlights the point that the models are given more freedom to deviate from our preconceptions.

It may seem natural to separate parametric and free-form models, but in my mind the fundamental distinction is whether models are *over*- or *under*-constrained. Over-constrained free-form models are actually akin to parametric models in two ways. First, there is some well-defined peak in the likelihood that we can explore to obtain useful conclusions with or without priors. Second, in general the only way to keep free-form models over-constrained is to limit the number of terms in Eq. (35).[8] The choice of basis functions and number of terms is not so different from the choice of parametric models.

The issues are different when free-form models are under-constrained. In general there will be many models that fit the data perfectly. To see this, consider a point-image lens and plug Eq. (35) into the constraint equations for the image positions and time delays:

$$\boldsymbol{x}_i^{\rm obs} = \sum_{\nu} a_\nu \, \nabla f_\nu \left( \boldsymbol{x}_i^{\rm obs} \right) + \boldsymbol{u}^{\rm mod}, \tag{36}$$

$$\frac{1}{2} \left( |\boldsymbol{x}_i|^2 - |\boldsymbol{x}_j|^2 \right) = \sum_{\nu} a_\nu \left[ f_\nu \left( \boldsymbol{x}_i^{\rm obs} \right) - f_\nu \left( \boldsymbol{x}_j^{\rm obs} \right) \right] + \left( \boldsymbol{x}_i^{\rm obs} - \boldsymbol{x}_j^{\rm obs} \right)^t \boldsymbol{u}^{\rm mod} + t_0^{-1} \Delta t_{ij}^{\rm obs}, \tag{37}$$

where $\Delta t_{ij}^{\rm obs} = t_j^{\rm obs} - t_i^{\rm obs}$ is the differential time delay between images $i$ and $j$. These equations are linear in the coefficients $a_\nu$, along with the source position $\boldsymbol{u}^{\rm mod}$ and time scale $t_0^{-1} \propto H_0$. When this system of equations is under-constrained it admits some linear solution space. Adding flux constraints may make things a little more complicated because the constraint equations may become non-linear, but the principles still hold. The key point is that all models in the solution space fit equally well, so the likelihood is uniform across the solution space, which means priors come to prominence as the only way to distinguish among the successful models.

[8] This is because fully free-form models have mainly been applied to lenses with point-like images, which provide a modest number of constraints. The exception is the work by Yoo et al. [42,43] discussed in Sect. 5.3.1.

There have been assorted approaches to free-form models that basically come down to different choices of basis functions and priors, and different data that lead to the models being over- or under-constrained.

### *5.3.1 Multipole models*

One approach to free-form models is to fix the radial profile of the lens galaxy but allow an arbitrary angular structure [28,42,43,76–79]. If we assume an isothermal profile, we can expand the galaxy potential in a multipole series to order $m_{\rm max}$,

$$\phi_{\rm gal}(r,\theta) = r\sum_{m=0}^{m_{\rm max}} (a_m \cos m\theta + b_m \sin m\theta)\,. \tag{38}$$

Note that the coefficient $b_0$ is irrelevant, while $a_1$ and $b_1$ correspond to a unobservable uniform deflection; so those three terms may be dropped from the expansion, leaving $2m_{\rm max} - 1$ coefficients to be determined. We may also want or need to include shear from the lens galaxy's environment following Eq. (32)

The full potential has the form Eq. (35), so a point-image lens yields position and time delay constraints that are linear in the multipole moments, $a_m$ and $b_m$, as well as the tidal shear amplitudes $\gamma_c$ and $\gamma_s$. A cute property of the isothermal profile is that the flux constraints are also linear in the multipole moments (but not the shear). The inverse magnification is

$$\begin{aligned}\mu^{-1} = 1 - \gamma_c^2 - \gamma_s^2 + \sum_{m=0}^{m_{\rm max}} \frac{m^2-1}{r}\,(a_m \cos m\theta + b_m \sin m\theta)\\ \times\,(1 - \gamma_c \cos 2\theta - \gamma_s \sin 2\theta)\,,\end{aligned} \tag{39}$$

so if we think of the flux ratio constraints as $F_i^{\rm obs}/F_j^{\rm obs} = \mu_i/\mu_j$ and rewrite that as $F_i^{\rm obs}\mu_i^{-1} = F_j^{\rm obs}\mu_j^{-1}$ we obtain equations that are linear in $a_m$ and $b_m$. The flux constraints are not linear in $\gamma_c$ and $\gamma_s$, but we can do a two-stage analysis in which we explicitly vary the non-linear parameters $(\gamma_c, \gamma_s)$ and at each step do a linear analysis for $\{a_m, b_m\}$.

When Evans and Witt [77] fit multipole models to lenses with anomalous flux ratios, they chose $m_{\rm max}$ so the number of unknown variables exactly matched the number of constraints. That seems odd, though, because the real galaxy has no way to "know" how many constraints we have measured. Congdon and Keeton [79] instead allowed $m_{\rm max}$ to be larger, which made the problem under-constrained. They imposed regularization by favoring the "minimum wiggle model" with the smallest deviations from elliptical symmetry. Congdon and Keeton found that even minimum wiggle models are still highly unrealistic for lenses with strong flux ratio anomalies, and concluded that multipoles do not provide a good explanation for flux ratio anomalies.

Yoo et al. [42,43] fit multipole models to Einstein rings modeled using the method from Sect. 4.3. The abundance of constraints from the rings made the models over-constrained, so Yoo et al. located the peak of the likelihood and used $\Delta\chi^2$ thresholds

to examine the range of allowed models. They found that Einstein rings strongly limit departures from elliptical symmetry in lens galaxy mass distributions.

### *5.3.2 Multipole/Taylor models*

Trotter et al. [80] (following a suggestion by Kochanek [28]) also used a multipole expansion for the angular structure of the potential, but allowed a more general radial dependence: they let the multipole coefficients be arbitrary functions of radius, $a_m(r)$ and $b_m(r)$, and expanded them in Taylor series around a radius $r_0$. The Einstein radius is a natural choice for the expansion radius. A general form of the expansion is

$$\phi(r,\theta) = \sum_{m=0}^{m_{\max}} \sum_{n=0}^{n_{\max}} \left(\frac{r}{r_0} - 1\right)^n (a_{mn} \cos m\theta + b_{mn} \sin m\theta) \,. \tag{40}$$

Not all terms are equally important, and Trotter et al. went to some effort to give physical interpretations to the terms and identify the ones that are most relevant for lens modeling. They retained a modest number of terms, and took advantage of extra position constraints from high-resolution VLBI observations that resolved the four images of MG J0414+0534 into multiple subcomponents [80,81], to keep the problem over-constrained. They used $\Delta\chi^2$ thresholds to quantify the parameter uncertainties. It would be interesting to embed multipole/Taylor models in a full Bayesian framework to enable a richer analysis of parameter uncertainties and covariances. That would also make it possible to include more terms in the expansion and constrain them with priors derived from the shapes of observed galaxies and simulated dark matter halos.

### *5.3.3 Pixelated mass maps*

Saha and Williams [82–84] have developed free-form models in which the basis functions correspond to mass pixels. Their PixeLens code applies to point-image lenses and uses position and time delay constraints, so the problem is fully linear (in the pixel densities, the source position, the time scale factor $t_0^{-1}$, and also the external shear amplitudes). With even modest spatial resolution in the mass map, the mass pixels outnumber the constraints so the problem is under-constrained. Saha and Williams deal with the resulting solution space in two ways. First, they impose priors to eliminate models that are grossly unphysical:

1. All pixel densities must be non-negative.
2. The density gradient must point within 45° of the lens center.
3. No pixel value may exceed the average of its neighbors by more than a factor of two (except for the central pixel).
4. The projected density profile must be steeper than $r^{-1/2}$.
5. If desired, the mass map may be required to have inversion symmetry.

These priors restrict the solution space but are otherwise non-informative, so the posterior is uniform over the allowed models. So the second step is to use Monte Carlo techniques to generate an ensemble of models drawn from the posterior. The ensemble

average is itself a successful model (since the constraints are linear, any linear combination of good models is also good) that gives a sense of the "typical" properties of the solutions. The statistics of the ensemble then characterize the range of successful models.

Saha, Williams and collaborators have used pixelated mass models for a variety of applications: examining the structure of lens mass distributions, comparing visible and dark matter distributions in lens galaxies, using measured time delays to constrain the Hubble constant, and predicting time delays for other lenses [85–91]. (They have also applied their method to cluster mass reconstructions, which are not discussed here.) It is worth noting that the priors are not very restrictive. In particular, requiring the surface mass density to be non-negative does not necessarily imply a non-negative 3-d density, let alone a non-negative distribution function (see [1]). Furthermore, the modeling analysis does not automatically exclude mass distributions that predict too many images. We should therefore consider pixelated mass models to provide the largest possible range of models that are (exactly) consistent with the data and have structure only on scales larger than the mass pixels; but we should not necessarily conclude that all solutions are physically plausible.

### 5.4 Hybrid models

In Sect. 5.3 we saw that free-form modeling has been used for lenses with point-like or elliptical sources. It has not been applied to lenses with general extended sources, because the source reconstruction method is non-linear in the lens parameters (see Sect. 4.5), and it is impractical to undertake a non-linear optimization of a large number of free-form modes. Instead, what I call "hybrid models" have been introduced to accommodate extended sources while still allowing a (quasi-)linear analysis of general lens potentials [38,39,55,56,92].

In hybrid models we obtain an arbitrary potential, $\phi = \phi_0 + \delta\phi$, as a combination of a reference model $\phi_0$ and perturbations $\delta\phi$.[9] The reference model is taken to have some restricted, parametric form, so it provides a modest number of non-linear parameters that must be searched explicitly. The perturbations, by contrast, are allowed to be fully free. The idea is that if the potential perturbations are small, we can do a Taylor series expansion and work to lowest order in $\delta\phi$, and thus linearize the problem. As we will see, there are several reasons why it may be necessary to iterate a few times, but at each step the analysis is linear.

Given a source model $S(\boldsymbol{u})$, we can predict the model image as

$$\begin{aligned} I^{\rm mod}(\boldsymbol{x}) &= S\left(\boldsymbol{x} - \boldsymbol{\alpha}_0 - \nabla\delta\phi\right), \\ &\approx S(\boldsymbol{x} - \boldsymbol{\alpha}_0) - \left[\nabla_u S(\boldsymbol{x} - \boldsymbol{\alpha}_0)\right]^t \left[\nabla_x \delta\phi(\boldsymbol{x})\right]. \end{aligned} \tag{41}$$

where $\boldsymbol{\alpha}_0 = \nabla\phi_0$ is the deflection from the reference model. In the second step we have taken the Taylor series expansion to first order in $\delta\phi$. Now let us write a pixelated

[9] Koopmans [55] discusses why hybrid models are formulated in terms of potential rather than mass perturbations.

version of Eq. (41), working with the pixelated source model, $\boldsymbol{s}$, as well as the pixelated potential perturbations, $\boldsymbol{\delta\phi}$. Following Sect. 4.5, the first term can be written as $\mathbf{L}_0\boldsymbol{s}$ where $\mathbf{L}_0$ is the lensing operator (for the reference model). The factor $\nabla_u S(\boldsymbol{x} - \boldsymbol{\alpha}_0)$ can be written as a matrix that is a function of the source model, $\mathbf{D}_s(\boldsymbol{s})$. If we use finite differencing to estimate the derivative of $\delta\phi(\boldsymbol{x})$, then we can write $\nabla_x \delta\phi(\boldsymbol{x})$ as $\mathbf{D}_\phi\,\boldsymbol{\delta\phi}$ where $\mathbf{D}_\phi$ is a finite difference matrix. (See [55] for explicit forms of the $\mathbf{D}_s(\boldsymbol{s})$ and $\mathbf{D}_\phi$ matrices.) Finally, we can account for PSF smearing by introducing an overall factor of the blurring operator $\mathbf{B}$. In total, then, we can write the pixelated model image as

$$\boldsymbol{d}^{\rm mod} = \mathbf{B}\left[\mathbf{L}_0\,\boldsymbol{s} - \mathbf{D}_s^t(\boldsymbol{s})\,\mathbf{D}_\phi\,\boldsymbol{\delta\phi}\right]. \tag{42}$$

Assuming we want to use quadratic regularization for both the source and potential reconstructions, we put

$$\begin{aligned}\chi^2 = &\left(\boldsymbol{d}^{\rm mod} - \boldsymbol{d}^{\rm obs}\right)^t \mathbf{S}_d^{-1} \left(\boldsymbol{d}^{\rm mod} - \boldsymbol{d}^{\rm obs}\right) \\ &+ \lambda_s\,\boldsymbol{s}^t\,\mathbf{H}_s^t\,\mathbf{H}_s\,\boldsymbol{s} \;+\; \lambda_\phi\,\boldsymbol{\delta\phi}^t\,\mathbf{H}_\phi^t\,\mathbf{H}_\phi\,\boldsymbol{\delta\phi},\end{aligned} \tag{43}$$

where $\mathbf{S}_d$ is the covariance matrix for the data. The first term quantifies the model's goodness of fit, while the second and third terms represent the regularization of the source and potential perturbations, respectively. Note that the source and potential regularizations can have independent forms and strengths.

In principle we want to solve for the source model, $\boldsymbol{s}$, and the potential perturbations, $\boldsymbol{\delta\phi}$, simultaneously. That is not a linear problem, however, because the last term in Eq. (42) involves a product of $\boldsymbol{s}$ and $\boldsymbol{\delta\phi}$. A way to deal with the non-linearity is to iterate as follows. The zeroth-order step is to optimize the reference model; this involves an explicit search of the non-linear parameters using methods discussed in Sect. 5.1. Then:

1. Set the potential perturbations to zero and reconstruct the source. With $\boldsymbol{\delta\phi} = 0$, the optimal value of $\boldsymbol{s}$ can be found by solving (q.v. Eq. 31)

$$\left(\mathbf{L}^t\,\mathbf{S}_d^{-1}\,\mathbf{L} + \lambda_s\,\mathbf{H}_s^t\,\mathbf{H}_s\right)\,\boldsymbol{s} = \mathbf{L}^t\,\mathbf{S}_d^{-1}\,\boldsymbol{d}^{\rm obs}, \tag{44}$$

   where $\mathbf{L} = \mathbf{B}\,\mathbf{L}_0$.
2. Hold the source fixed and compute the potential perturbations. With $\boldsymbol{s}$ fixed, the optimal value for $\boldsymbol{\delta\phi}$ can be found by solving

$$\left(\mathbf{A}^t\,\mathbf{S}_d^{-1}\,\mathbf{A} + \lambda_\phi\,\mathbf{H}_\phi^t\,\mathbf{H}_\phi\right)\,\boldsymbol{\delta\phi} = \mathbf{A}^t\,\mathbf{S}_d^{-1}\left(\mathbf{L}\,\boldsymbol{s} - \boldsymbol{d}^{\rm obs}\right), \tag{45}$$

   where $\mathbf{A} = \mathbf{B}\,\mathbf{D}_s^t(\boldsymbol{s})\,\mathbf{D}_\phi$.
3. Use the potential perturbations to update the reference model: $\phi_0^{\rm new} = \phi_0^{\rm old} + \delta\phi$. Taking this as the new reference model, return to step #1 and repeat until converged.

Another benefit of this process is that at early stages the potential perturbations may not be small, but as the iteration proceeds the perturbations should shrink and validate the Taylor series approximation.

Note from Eq. (44) that the source model is driven directly by the observed image (the factor of $\boldsymbol{d}^{\rm obs}$ on the right-hand side). The potential perturbations, by contrast, are driven by the image residuals of the reference model (the factor of $\mathbf{L}\,\boldsymbol{s} - \boldsymbol{d}^{\rm obs}$ on the right-hand side of Eq. 45).

The full analysis depends, of course, on the regularizations ($\lambda_s$, $\mathbf{H}_s$, $\lambda_\phi$, and $\mathbf{H}_\phi$). In the Bayesian framework, we can use the Bayesian evidence at a second level of inference to determine the optimal regularization parameters $\lambda_s$ and $\lambda_\phi$, and to compare different regularization schemes (see [54,56] for details).

Suyu et al. [92,93] have used hybrid models to handle the complexity in B1608+656 associated with having two galaxies inside the Einstein radius. Suyu et al. started with a parametric two-galaxy model by Koopmans et al. [41] and then refined it using pixelated potential reconstruction. (They also accounted for many other details including light from the lens galaxies, and extinction of the lensed images by dust in the lens galaxies.) Their joint Bayesian analysis of constraints from lensing and WMAP supports a wide range of conclusions involving the mass distribution of the lens galaxies and their environment, the Hubble constant, and curvature and dark energy parameters. Koopmans, Vegetti and collaborators [55,56,94,95] have used hybrid models as the basis of a method to search for unseen mass clumps in lens galaxies and place constraints on dark matter substructure. A mass clump near the Einstein radius will distort an extended image in a characteristic way that can be uncovered by simultaneously reconstructing the (pixelated) source and potential. Vegetti et al. [95] reported the detection of a mass clump in the SLACS lens SDSS J0946+1006, and used parametric models to infer a clump mass of $(3.51 \pm 0.15) \times 10^6\, M_\odot$.

### 5.5 Inference without modeling

The discussion so far has focused on bona fide modeling, in which we attempt to fit observed lens systems in as much detail as possible. Now consider a different approach to inference that gives more explicit weight to priors and statistical arguments. I mentioned in Sect. 4.2 a treatment of data in which we characterize a lens with minimal information and then define a likelihood that is uniform among models that match the information (and zero otherwise). The goal is to find a very broad range of models that are consistent with an observed lens configuration.

One motivation is to test certain relations among lens flux ratios and time delays that are predicted to be universal for smooth mass distributions [33,35,96]. Examining a large set of models allows us to verify that the relations are (mostly) model-independent. The predicted relations do depend on the image configuration, though, so we want to focus on models that are broadly consistent with observed lenses.

A second motivation is to build independent astrophysical knowledge into the analysis, through priors. If we use *objective* priors—motivated by or even taken from observations of real galaxies—then we can identify the full range of lens properties that can be produced by "realistic" mass distributions. (The degree of realism depends,

of course, on the quality of the priors.) This analysis is conceptually related to under-constrained free-form modeling (q.v. Sect. 5.3.3), but it yields a much larger solution space because we attempt to match less information, and it purposely uses priors that are informative.

Keeton et al. [33,35] and Congdon et al. [36] used this method to predict the range of flux ratios and time delays that can be produced by smooth mass distributions. For priors they used measurements of ellipticities and $m = 4$ multipole moments from samples of early-type galaxies, and tidal shear distributions from $N$-body simulations. Their goal was to look for outliers—lenses that are not consistent with smooth mass distributions and thus are candidates for containing dark matter substructure or other interesting complexity in the lens potential. Oguri [34] used a similar method to perform a statistical analysis of lensing constraints on the Hubble constant.

**Acknowledgments** Support for this work has been provided by NSF through grant AST-0747311.

## References

1. Kochanek, C.S.: In: Meylan, G., Jetzer, P., North, P. (eds.) Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, pp. 91–268 (2006)
2. Courbin F., Saha P., Schechter P.L.: In: Courbin, F., Minniti, D. (eds.) Gravitational Lensing: An Astrophysical Tool, Lecture Notes in Physics, vol. 608, pp. 1–54. Springer-Verlag, Berlin (2002)
3. Courbin, F.: Preprint. arXiv:astro-ph/0304497 (2003)
4. Schechter, P.L.: In: Mellier, Y., Meylan, G. (eds.) Gravitational Lensing Impact on Cosmology, IAU Symposium, vol. 225, pp. 281–296 (2005). doi:10.1017/S1743921305002085
5. Koopmans, L.V.E.: In: Davies, J., Disney, M. (eds.) IAU Symposium, vol. 244, pp. 196–205 (2008). doi:10.1017/S1743921307013993
6. Bartelmann, M.: Applications of Gravitational Lensing in Cosmology. pp. 213–256. Springer-Verlag, Berlin (2006)
7. Kneib, J.: In: Plionis, M., López-Cruz, O., Hughes, D. (eds.) A Pan-Chromatic View of Clusters of Galaxies and the Large-Scale Structure, Lecture Notes in Physics, vol. 740, pp. 213–253. Springer-Verlag, Berlin (2008)
8. Schneider, P., Ehlers, J., Falco, E.E.: Gravitational Lenses. Springer-Verlag, Berlin (1992)
9. Petters, A.O., Levine, H., Wambsganss, J.: Singularity Theory and Gravitational Lensing. Birkhäuser, Boston (2001)
10. Schramm, T.: Astron. Astrophys. **231**, 19 (1990)
11. Keeton, C.R.: Preprint. arXiv:astro-ph/0102341 (2001)
12. van de Ven, G., Mandelbaum, R., Keeton, C.R.: Mon. Not. R. Astron. Soc. **398**, 607 (2009). doi:10.1111/j.1365-2966.2009.15167.x
13. Gelman, A.B., Carlin, J.S., Stern, H.S., Rubin, D.B.: Bayesian Data Analysis. Chapman & Hall/CRC, Boca Raton (1995)
14. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C: The Art of Scientific Computing, 2nd edn. Cambridge University Press, Cambridge (1992)
15. Skilling, J.: In: Fischer, R., Preuss, R., Toussaint, U.V. (eds.) American Institute of Physics Conference Series, vol. 735, pp. 395–405 (2004). doi:10.1063/1.1835238
16. Mukherjee, P., Parkinson, D., Liddle, A.R.: Astrophys. J. Lett. **638**, L51 (2006). doi:10.1086/501068
17. Rhie, S.H.: Preprint. arXiv:astro-ph/0103463 (2001)
18. Khavinson, D., Neumann, G.: Proc AMS **134**, 1077 (2006)
19. An, J.H., Evans, N.W.: Mon. Not. R. Astron. Soc. **369**, 317 (2006). doi:10.1111/j.1365-2966.2006.10303.x
20. Fassnacht, C.D., Keeton, C.R., Khavinson, D.: Preprint. arXiv:0708.2684 (2007)
21. Khavinson,D., Lundberg, E.: Preprint. arXiv:0908.3310 (2009)
22. Bergweiler, W., Eremenko, A.: Preprint. arXiv:0908.4595 (2009)
23. Blandford, R.D., Kochanek, C.S.: Astrophys. J. **321**, 658 (1987). doi:10.1086/165660

24. Kochanek, C.S., Blandford, R.D.: Astrophys. J. **321**, 676 (1987). doi:10.1086/165661
25. Keeton, C.R.: Preprint. arXiv:astro-ph/0102340 (2001)
26. Shewchuk, J.R.: In: Lin, M.C., Manocha, D. (eds.) Applied Computational Geometry: Towards Geometric Engineering, Lecture Notes in Computer Science, vol. 1148, pp. 203–222. From the First ACM Workshop on Applied Computational Geometry. Springer-Verlag, Berlin (1996)
27. Shewchuk, J.R.: Comput. Geometry **22**(1–3), 21 (2002). doi:10.1016/S0925-7721(01)00047-5
28. Kochanek, C.S.: Astrophys. J. **373**, 354 (1991). doi:10.1086/170057
29. Saha, P., Williams, L.L.R.: Astron. J. **125**, 2769 (2003). doi:10.1086/375204
30. Linder, E.V.: Phys. Rev. D **70**(4), 043534 (2004). doi:10.1103/PhysRevD.70.043534
31. Coe, D., Moustakas, L.A.: Astrophys. J. **706**, 45 (2009). doi:10.1088/0004-637X/706/1/45
32. LSST Science Collaborations: LSST Science Book, Version 2.0. (LSST Corporation, Tucson 2009)
33. Keeton, C.R., Gaudi, B.S., Petters, A.O.: Astrophys. J. **598**, 138 (2003). doi:10.1086/378934
34. Oguri, M.: Astrophys. J. **660**, 1 (2007). doi:10.1086/513093
35. Keeton, C.R., Gaudi, B.S., Petters, A.O.: Astrophys. J. **635**, 35 (2005). doi:10.1086/497324
36. Congdon, A.B., Keeton, C.R., Nordgren, C.E.: Astrophys. J. **709**, 552 (2010)
37. Williams, L.L.R., Foley, P., Farnsworth, D., Belter, J.: Astrophys. J. **685**, 725 (2008). doi:10.1086/591227
38. Blandford, R., Surpi, G., Kundić, T.: In: Brainerd, T.G., Kochanek, C.S. (eds.) Gravitational Lensing: Recent Progress and Future Goals, Astronomical Society of the Pacific Conference Series, vol. 237, pp. 65–74 (2001)
39. Suyu, S.H., Blandford, R.D.: Mon. Not. R. Astron. Soc. **366**, 39 (2006). doi:10.1111/j.1365-2966.2005.09854.x
40. Kochanek, C.S., Keeton, C.R., McLeod, B.A.: Astrophys. J. **547**, 50 (2001). doi:10.1086/318350
41. Koopmans, L.V.E., Treu, T., Fassnacht, C.D., Blandford, R.D., Surpi, G.: Astrophys. J. **599**, 70 (2003). doi:10.1086/379226
42. Yoo, J., Kochanek, C.S., Falco, E.E., McLeod, B.A.: Astrophys. J. **626**, 51 (2005). doi:10.1086/429959
43. Yoo, J., Kochanek, C.S., Falco, E.E., McLeod, B.A.: Astrophys. J. **642**, 22 (2006). doi:10.1086/500968
44. Kochanek, C.S., Narayan, R.: Astrophys. J. **401**, 461 (1992). doi:10.1086/172078
45. Ellithorpe, J.D., Kochanek, C.S., Hewitt, J.N.: Astrophys. J. **464**, 556 (1996). doi:10.1086/177346
46. Wucknitz, O.: Mon. Not. R. Astron. Soc. **349**, 1 (2004). doi:10.1111/j.1365-2966.2004.07513.x
47. Wallington, S., Kochanek, C.S., Koo, D.C.: Astrophys. J. **441**, 58 (1995). doi:10.1086/175335
48. Wallington, S., Narayan, R., Kochanek, C.S.: Astrophys. J. **426**, 60 (1994). doi:10.1086/174039
49. Wallington, S., Kochanek, C.S., Narayan, R.: Astrophys. J. **465**, 64 (1996). doi:10.1086/177401
50. Wayth, R.B., Webster, R.L.: Mon. Not. R. Astron. Soc. **372**, 1187 (2006). doi:10.1111/j.1365-2966.2006.10922.x
51. Warren, S.J., Dye, S.: Astrophys. J. **590**, 673 (2003). doi:10.1086/375132
52. Dye, S., Warren, S.J.: Astrophys. J. **623**, 31 (2005). doi:10.1086/428340
53. Treu, T., Koopmans, L.V.E.: Astrophys. J. **611**, 739 (2004). doi:10.1086/422245
54. Suyu, S.H., Marshall, P.J., Hobson, M.P., Blandford, R.D.: Mon. Not. R. Astron. Soc. **371**, 983 (2006). doi:10.1111/j.1365-2966.2006.10733.x
55. Koopmans, L.V.E.: Mon. Not. R. Astron. Soc. **363**, 1136 (2005). doi:10.1111/j.1365-2966.2005.09523.x
56. Vegetti, S., Koopmans, L.V.E.: Mon. Not. R. Astron. Soc. **392**, 945 (2009). doi:10.1111/j.1365-2966.2008.14005.x
57. Keeton, C.R., Kochanek, C.S.: Astrophys. J. **487**, 42 (1997). doi:10.1086/304583
58. Saha, P.: Astron. J. **120**, 1654 (2000). doi:10.1086/301581
59. Kochanek, C.S.: Astrophys. J. **578**, 25 (2002). doi:10.1086/342476
60. Cohn, J.D., Kochanek, C.S., McLeod, B.A., Keeton, C.R.: Astrophys. J. **554**, 1216 (2001). doi:10.1086/321412
61. Kochanek, C.S., Morgan, N.D., Falco, E.E., McLeod, B.A., Winn, J.N., Dembicky, J., Ketzeback, B.: Astrophys. J. **640**, 47 (2006). doi:10.1086/499766
62. Koopmans, L.V.E., Bolton, A., Treu, T., Czoske, O., Auger, M.W., Barnabè, M., Vegetti, S., Gavazzi, R., Moustakas, L.A., Burles, S.: Astrophys. J. Lett. **703**, L51 (2009). doi:10.1088/0004-637X/703/1/L51
63. Fadely, R., Keeton, C.R., Nakajima, R., Bernstein, G.M.: Astrophys. J. **711**, 246 (2010)
64. Keeton, C.R., Kochanek, C.S., Seljak, U.: Astrophys. J. **482**, 604 (1997). doi:10.1086/304172
65. Mao, S., Schneider, P.: Mon. Not. R. Astron. Soc. **295**, 587 (1998). doi:10.1046/j.1365-8711.1998.01319.x

66. Metcalf, R.B., Madau, P.: Astrophys. J. **563**, 9 (2001). doi:10.1086/323695
67. Chiba, M.: Astrophys. J. **565**, 17 (2002). doi:10.1086/324493
68. Dalal, N., Kochanek, C.S.: Astrophys. J. **572**, 25 (2002). doi:10.1086/340303
69. Momcheva, I., Williams, K., Keeton, C., Zabludoff, A.: Astrophys. J. **641**, 169 (2006). doi:10.1086/500382
70. Fassnacht, C.D., Gal, R.R., Lubin, L.M., McKean, J.P., Squires, G.K., Readhead, A.C.S.: Astrophys. J. **642**, 30 (2006). doi:10.1086/500927
71. Falco, E.E., Gorenstein, M.V., Shapiro, I.I.: Astrophys. J. Lett. **289**, L1 (1985). doi:10.1086/184422
72. Gorenstein, M.V., Shapiro, I.I., Falco, E.E.: Astrophys. J. **327**, 693 (1988). doi:10.1086/166226
73. Kochanek, C.S.: Astrophys. J. **382**, 58 (1991). doi:10.1086/170693
74. Keeton, C.R., Zabludoff, A.I.: Astrophys. J. **612**, 660 (2004). doi:10.1086/422745
75. Koopmans, L.V.E., Treu, T., Bolton, A.S., Burles, S., Moustakas, L.A.: Astrophys. J. **649**, 599 (2006). doi:10.1086/505696
76. Evans, N.W., Witt, H.J.: Mon. Not. R. Astron. Soc. **327**, 1260 (2001). doi:10.1046/j.1365-8711.2001.04826.x
77. Evans, N.W., Witt, H.J.: Mon. Not. R. Astron. Soc. **345**, 1351 (2003). doi:10.1046/j.1365-2966.2003.07057.x
78. Kawano, Y., Oguri, M., Matsubara, T., Ikeuchi, S.: Publ. Astron. Soc. Jpn. **56**, 253 (2004)
79. Congdon, A.B., Keeton, C.R.: Mon. Not. R. Astron. Soc. **364**, 1459 (2005). doi:10.1111/j.1365-2966.2005.09699.x
80. Trotter, C.S., Winn, J.N., Hewitt, J.N.: Astrophys. J. **535**, 671 (2000). doi:10.1086/308861
81. Ros, E., Guirado, J.C., Marcaide, J.M., Pérez-Torres, M.A., Falco, E.E., Muñoz, J.A., Alberdi, A., Lara, L.: Astron. Astrophys. **362**, 845 (2000)
82. Saha, P., Williams, L.L.R.: Mon. Not. R. Astron. Soc. **292**, 148 (1997)
83. Williams, L.L.R., Saha, P.: Astron. J. **119**, 439 (2000). doi:10.1086/301234
84. Saha, P., Williams, L.L.R.: Astron. J. **127**, 2604 (2004). doi:10.1086/383544
85. Williams, L.L.R., Saha, P.: Astron. J. **128**, 2631 (2004). doi:10.1086/426007
86. Ferreras, I., Saha, P., Williams, L.L.R.: Astrophys. J. Lett. **623**, L5 (2005). doi:10.1086/429995
87. Saha, P., Coles, J., Macciò, A.V., Williams, L.L.R.: Astrophys. J. Lett. **650**, L17 (2006). doi:10.1086/507583
88. Saha, P., Courbin, F., Sluse, D., Dye, S., Meylan, G.: Astron. Astrophys. **450**, 461 (2006). doi:10.1051/0004-6361:20052929
89. Saha, P., Williams, L.L.R., Ferreras, I.: Astrophys. J. **663**, 29 (2007). doi:10.1086/518083
90. Read, J.I., Saha, P., Macciò, A.V.: Astrophys. J. **667**, 645 (2007). doi:10.1086/520714
91. Ferreras, I., Saha, P., Burles, S.: Mon. Not. R. Astron. Soc. **383**, 857 (2008). doi:10.1111/j.1365-2966.2007.12606.x
92. Suyu, S.H., Marshall, P.J., Blandford, R.D., Fassnacht, C.D., Koopmans, L.V.E., McKean, J.P., Treu, T.: Astrophys. J. **691**, 277 (2009). doi:10.1088/0004-637X/691/1/277
93. Suyu, S.H., Marshall, P.J., Auger, M.W., Hilbert, S., Blandford, R.D., Koopmans, L.V.E., Fassnacht, C.D., Treu, T.: Astrophys. J. **711**, 201 (2010)
94. Vegetti, S., Koopmans, L.V.E.: Mon. Not. R. Astron. Soc. **400**, 1583 (2009). doi:10.1111/j.1365-2966.2009.15559.x
95. Vegetti, S., Koopmans, L.V.E., Bolton, A., Treu, T., Gavazzi, R.: Preprint. arXiv:0910.0760 (2009)
96. Congdon, A.B., Keeton, C.R., Nordgren, C.E.: Mon. Not. R. Astron. Soc. **389**, 398 (2008). doi:10.1111/j.1365-2966.2008.13604.x



View publication stats