

Joint Semi-Analytic Multipole Priors from Galaxy Isophotes and Constraints from Lensed Arcs

Maverick S. H. Oh ,^{a,1} Anna Nierenberg ,^a Daniel Gilman ,^{b,c} Simon Birrer 

^aUniversity of California, Merced,
5200 North Lake Rd. Merced, CA 95343, USA

^bThe University of Chicago,
5801 S. Ellis Ave. Chicago, IL 60637, USA

^cBrinson Prize Fellow

^dStony Brook University,
100 Nicolls Rd, Stony Brook, NY 11794, USA

E-mail: soh39@ucmerced.edu

Abstract. Flux-ratio anomalies in quadruply imaged quasars are sensitive to the imprint of low-mass dark-matter haloes. The reliability of detection depends on the robustness of the smooth mass model. Optical surveys show that massive early-type galaxies similar to galaxy-scale gravitational lenses depart from perfect ellipticity, exhibiting $m = 3$ and $m = 4$ multipole distortions. We construct the semi-analytic, five-dimensional joint population prior for the $m = 3$ and $m = 4$ amplitude and orientation as well as the axis ratio of the deflector, calibrated on the sample of 840 SDSS E/S0 galaxies. The parameters are fitted via hierarchical Bayesian modeling, minimizing a joint Jensen–Shannon divergence between model and data. We use this prior to model the mass distribution of mock lenses with HST quality data with different multipole amplitudes. We find that we robustly measure the true multipole amplitudes and orientations. Compared to fits that use only the four point-image positions, adding the lensed host-galaxy arcs tightens the 68 % credible regions of multipole parameters by factors of 3–12 and reduces the predicted flux-ratio uncertainties by a mean factor of ~ 6 . This analysis does not include substructure or a complex source, and thus can be considered an upper limit on the expected improvement. The combination of arc information and realistic multipole priors therefore yields an order-of-magnitude improvement in smooth mass model precision, paving the way for more robust measurements of dark-matter substructure.

¹Corresponding author.

Contents

1	Introduction	1
2	Quadruply lensed quasar modeling	3
2.1	Mass model	3
2.1.1	Elliptical base mass model	3
2.1.2	Multipole perturbations	3
2.1.3	External shear	5
2.2	Light components	5
3	Population prior for multipole parameter values	5
4	Mock–data construction	11
4.1	Parameter setting	11
4.2	Data quality	12
5	Inference framework	12
5.1	Likelihood and priors	12
5.2	Sampling strategy	13
5.3	Derived quantities	13
6	Results	14
7	Discussion and conclusion	14
A	Comparison of multipole conventions	20
B	Optimization Details	20
C	Generation of the observational probability densities	21
D	Modeling details	22
E	Ellipticity parameterization and prior	22

1 Introduction

Dark matter dominates the mass budget of the Universe, yet its particle nature remains unknown [1]. Strong gravitational lensing of quasars provides one of the few probes of dark matter on sub-galactic scales because the deflection angles are set directly by the mass distribution—even in the absence of luminous tracers [2]. While various methods such as stellar stream perturbations and galaxy kinematics use baryonic tracers to infer the presence of low-mass dark matter halos [e.g. 3, 4], these approaches are often subject to degeneracies with baryonic physics. In contrast, gravitational lensing offers a more direct and clean probe of the mass distribution, enabling discrimination among dark matter models based on the abundance and internal structure of subhalos [see 2, and references therein].

The flux ratios of multiple quasar images are among the key observables of strong gravitational lensing. Flux ratio anomalies refer to the disparity between the observed flux ratios of lensed images and the flux ratios expected from a smooth mass distribution, which represents the large-scale distribution of the mass of the main lensing galaxy and its dark matter halo (a.k.a. “macromodel”). The macromodel is primarily responsible for causing multiple images of the background source to appear and determining their positions. Low-mass dark matter halos within the lensing galaxy and along the line of sight (a.k.a. “substructure”) make relatively little impact to the image positions relative to measurement uncertainties, but can introduce perturbations to the lensed image magnifications [see 2, 5–7, and references therein]. Note that such flux ratio anomalies may also arise from microlensing [e.g. 8] or inaccuracies in the smooth lens model [e.g. 9]. These effects need to be considered carefully when attributing flux ratio anomalies to dark matter substructure. Studies on multiple lens systems have shown that the amplitude and frequency of flux ratio anomalies exceed what can be accounted for by these effects alone, lending support to the presence of low-mass dark matter halos [e.g. 6, 10, 11].

Flux ratio anomalies can reveal the presence of low-mass substructures and have been used to infer population level statistics of their mass function and mass distribution [12–14]. However, robust inference from flux ratios requires an accurate smooth mass model. Traditional analyses adopt an Elliptical Power Law (EPL) profile plus external shear [e.g. 15–18], while recent studies [e.g. 19–21] have begun investigating the effect of deviations from ellipticity as detected in optical surveys of lens-like elliptical galaxies [22, Hao et al. 2006, hereafter H06].

Multipoles describe higher order perturbations to the mass distribution that cannot be captured by an elliptical profile [7, 11, 21, 23–28]. Of particular interest are multipoles of order $m = 3$ and $m = 4$ as these are prominent deviations from ellipticity observed in the light distribution of field elliptical galaxies [22]. However, *mass* multipoles have not been well measured compared to *light* multipoles, and there has not been empirically motivated priors on the multipole amplitudes ($a_3/a, a_4/a$) and orientations ($\phi_3 - \phi_0, \phi_4 - \phi_0$) as a readily usable form so far. Without such priors, estimating the correct multipole parameters from observed data can be challenging due to degeneracy. [29] pointed out that, when multipoles are not included in the lens modeling, they may appear in the external shear that does not agree with an independently measured cosmic shear.

Existing constraints on dark matter from [30–33] include an $m = 4$ multipole term, which adds boxyness and diskyness to the main deflector mass profile. The amplitude of this mass component is constrained by image positions and flux ratios jointly with the substructure properties. A recent simulation work presented the dark matter analysis pipeline with both $m = 3$ and $m = 4$ multipoles [20]. A warm dark matter constraint using JWST MIRI observations with both $m = 3$ and $m = 4$ multipoles and flexible ϕ_3, ϕ_4 was also presented recently [34]. These works adopted simple Gaussian priors for the multipole distribution. We build on these works by increasing the complexity of the prior to be both more flexible, as well as conditionally dependent on ellipticity. This work provides a continuous and differentiable joint distribution of multipole parameters to be used as a realistic prior.

Extended arcs—the lensed host-galaxy light—offer a route to break the difficulty of measuring multipoles. Because arcs wrap around the deflector, they probe the lensing potential over a much larger azimuthal range than the four point images alone [18, 35]. In this paper, we demonstrate that incorporating arc information, together with a population prior derived from H06, tightens the posterior on multipole parameters and predicted flux ratios.

Traditionally, only the lensed point source positions have been used to constrain the smooth-mass distribution. This yields large uncertainties of order 10-50% in the underlying smooth-model flux ratios which are comparable to or larger than the measurement uncertainties of the actual flux ratios themselves [see, e.g. 36]. Existing measurements of lens flux ratios from HST reach, on average, 6% precision [36], while mid-IR flux ratios measured with JWST can reach precisions of 1% [37]. Improving the precision of the smooth-model flux ratio predictions would therefore make a significant impact on the constraining power of gravitational lenses.

The remainder of this paper is organized as follows. Section 2 introduces our mass- and light-model parameterization, including the multipoles. Section 3 derives the population prior from the H06 catalog. Section 4 describes the mock-data generation and observational setups. Our inference framework is detailed in Section 5, the results are presented in Section 6, and Section 7 discusses implications for future dark-matter studies. Throughout we adopt $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. LENSTRONOMY² [38, 39] is used for lens modeling and fitting processes. The multipole prior is released as open-source Python software MULTIPOLEPRIOR³, along with additional scripts used for its development and visualization for transparency and reproducibility.

2 Quadruply lensed quasar modeling

In this section, we describe the model we use to describe quadruply lensed quasars and their host galaxies.

2.1 Mass model

The mass model includes three components: elliptical base mass model, multipole perturbations, and shear. Each of them are detailed below. Note that dark matter substructures are not included in this work to clearly demonstrate the impact of multipoles and lensed arcs without substructure lensing.

2.1.1 Elliptical base mass model

We adopt Elliptical power-law (EPL), of which the projected surface-mass density takes the form $\kappa(x, y) = \frac{3-\gamma}{2} \left(\frac{\theta_E}{\sqrt{qx^2+y^2/q}} \right)^{\gamma-1}$ when its major axis is aligned with the x -axis, where γ is the 3-D logarithmic slope (2 for isothermal), θ_E the circularized Einstein radius, and $q=b/a$ the projected axis ratio [40]. In general, the orientation of the ellipse is denoted by ϕ_0 , the angle of the major axis and x -axis.

2.1.2 Multipole perturbations

Multipole mass profiles add azimuthal perturbation to the mass profile of the lens, on top of the elliptical profile. We use the circular multipole convention [9, 26] to be consistent with H06. However, it should be noted that the elliptical multipoles have been shown to produce more physically meaningful optical and mass distributions at high ellipticity [see 41, 42] and are therefore recommended for future optical surveys.

When an order- m multipole is added to an elliptical profile, the isodensity (i.e. constant convergence κ) contours are deformed from a purely elliptical isodensity contours. The

²<https://github.com/lenstronomy/lenstronomy>

³<https://github.com/Maverick-Oh/multipoleprior>

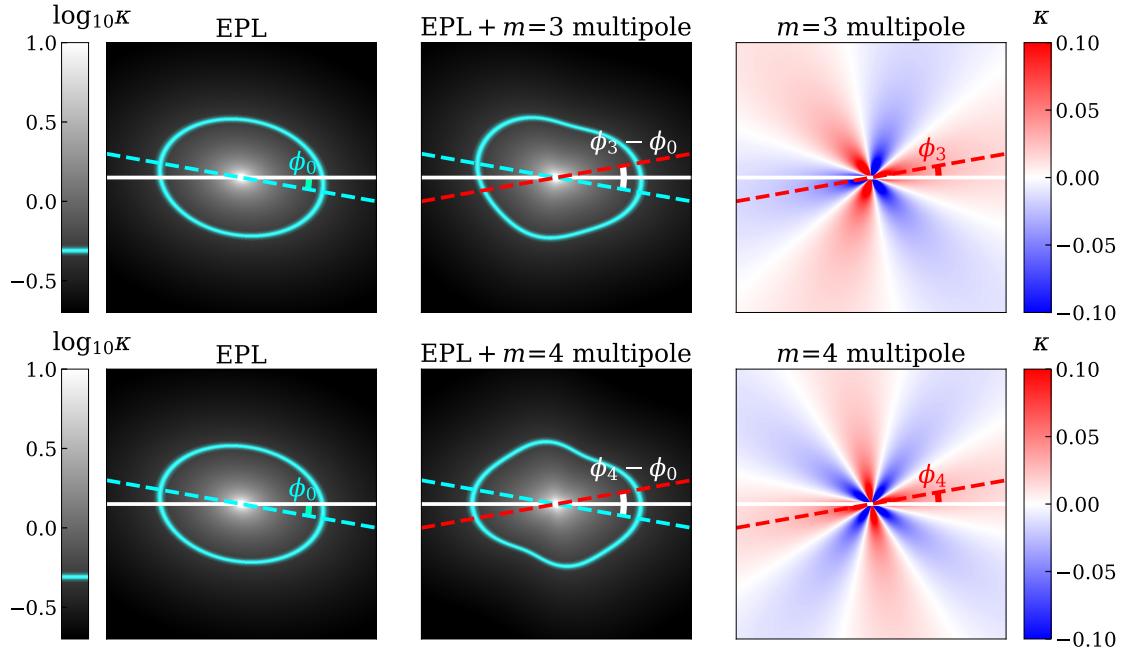


Figure 1. Impact of *misaligned* ($\phi_m \neq \phi_0$) $m=3$ (**upper**) and $m=4$ (**lower**) multipole profile on the convergence with the angular conventions ϕ_0 and ϕ_m . The isodensity curve is shown in cyan. (**Left**) Convergence of EPL-only lens mass model with $\phi_0 = -0.175$ (-10°). (**Middle**) Convergence of EPL+multipole. (**Right**) Convergence of $m=3$ and $m=4$ multipole with $a_m/a = 0.05$ and $\phi_m = 0.175$ (10°). For better visualization, $|a_m/a|$ is set to be larger than expected in typical systems.

deviation of the new isodensity contour from the ellipse can be expressed with a cosine function $\delta r = a_m \cos(m(\phi - \phi_m))$, where r is the polar radius $r = \sqrt{x^2 + y^2}$ and ϕ_m is the angle of the multipole profile's orientation ϕ_m . Note that the deviation amplitude a_m depends on which contour is chosen. When geometric similarity is assumed between the contours on different scales, the radial deviation is proportional to the size of the ellipse; i.e. $a_m \propto a$, where a is the semi-major axis of the ellipse and satisfies $a = \frac{1}{2\kappa} \theta_E / \sqrt{q}$. The ratio a_m/a is the key parameter that determines the shape of the deformed isodensity contour.

Choosing the standard isodensity contour with $\kappa = \frac{1}{2}$ gives the semi-major axis of the ellipse as the effective Einstein radius $a(\kappa = \frac{1}{2}) = \theta_E / \sqrt{q}$. The deviation amplitude there, $a_m(\kappa = \frac{1}{2})$, is used for setting the deflection potential. The deviation of the isodensity contours from the standard ellipse with $\kappa = \frac{1}{2}$ is

$$\delta r|_{\kappa=\frac{1}{2}} = a_m(\kappa=\frac{1}{2}) \cos(m(\phi - \phi_m)). \quad (2.1)$$

With a given value of a_m/a , the $a_m(\kappa = \frac{1}{2})$ value is calculated by

$$a_m(\kappa = \frac{1}{2}) = a_m/a \quad a(\kappa = \frac{1}{2}) = a_m/a \frac{\theta_E}{\sqrt{q}}. \quad (2.2)$$

Thus, the deflection potential can be set up based on either a_m/a or $a_m(\kappa = \frac{1}{2})$. In this paper, we stick to a_m/a convention.

We are interested in studying $m = 3$ and $m = 4$ multipoles. The $m = 3$ and $m = 4$ multipoles measure the triangle-like and quadrangle-like deformation on the isodensity

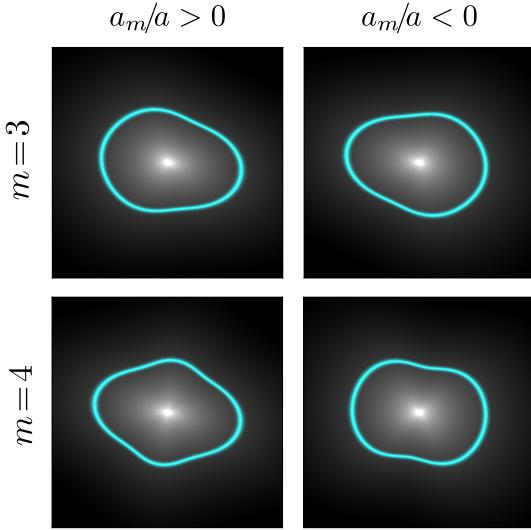


Figure 2. Impact of *aligned* ($\phi_m = \phi_0$) $m = 3$ (upper) and $m = 4$ (lower) multipole profile on the convergence with different signs of a_m/a . **(Left)** Convergence of EPL+multipole with $a_m/a = +0.05$. **(Right)** Convergence of EPL+multipole with $a_m/a = -0.05$. For better visualization, $|a_m/a|$ is set to be larger than expected in typical systems. An isodensity contour is highlighted in cyan color.

contour, respectively. They are parameterized with their multipole strength a_m/a and their angle relative to the elliptical profile $\phi_m - \phi_0$, where ϕ_0 refers to the angle of the elliptical profile (see Figure 1).

The $m = 4$ multipole is also known as boxy/diskyness, because it makes an elliptical profile either boxy or disk-like. When the orientation of the multipole profile and the elliptical profile are well aligned (i.e. $\phi_m \approx \phi_0$, where ϕ_0 is the angle of the elliptical profile), $a_4/a > 0$ results in a disk-like profile and $a_4/a < 0$ a boxy profile. Meanwhile, the sign of a_3/a changes only the orientation of the deformation, not its overall shape (see Figure 2).

Note that there are different conventions on how to define multipole variables and we follow the convention of [9, 28, 43]. We allow both signs of a_m and restrict $\phi_m \in (-\pi/2m, \pi/2m]$ to avoid angular degeneracy. In Appendix A we provide relations between this and other commonly used conventions.

2.1.3 External shear

We include external shear which can be caused by external sources such as galaxy clusters or large-scale structure [44–46].

2.2 Light components

The deflector light and the quasar-host arcs are each represented by an elliptical Sérsic profile [47]. Deflected quasar images are treated as point sources (delta functions convolved with the PSF).

3 Population prior for multipole parameter values

In this section we explore what reasonable priors are for the mass multipole parameters. The multipole parameters of the lens mass profiles have not been well constrained from the

observation so far. However, *optical* surveys of elliptical-galaxy light profiles can serve as empirical benchmarks and place conservative upper limits on multipole amplitudes, because (i) dark-matter halos are well described by simple ellipsoids with negligible higher-order structure [see 48, 49, and references therein]; (ii) the baryonic disc and bulge introduce non-elliptical perturbations [e.g. 19]; and (iii) baryons contribute only $\sim 30\text{--}40\%$ of the projected mass within the Einstein radius [50, 51]. The optical a_m and a are defined the similarly as the mass profile's isodensity contours, but instead using isophote contours. The ratio of deformation amplitude and the semi-major axis, a_m/a , has been measured with different galaxies and at different radii [21, 22, 52–55].

H06 has shown that $m = 3$ and $m = 4$ multipoles have different distributions⁴. However, plotting a_m/a together with $\phi_m - \phi_0$ with different q ranges reveals more correlations from the data of H06. See “Observation” columns of Figure 3 for the distribution of $(a_3/a, \phi_3 - \phi_0)$ and $(a_4/a, \phi_4 - \phi_0)$ with different q ranges from the isophotes of 840 E/S0 galaxies⁵, with the fourth power of velocity dispersion v_{dis}^4 as weight⁶, of which the original data is provided by H06. Four important correlations can be observed:

1. a_4/a tends to be positive and greater (more disk-like) when q is smaller (more elliptical).
2. $\phi_4 - \phi_0$ tends to be close to 0 (more aligned) when a_4/a is positive & greater (more disk-like).
3. a_3/a is symmetric around 0 regardless of q ; whereas the width of the true distribution varies.
4. $\phi_3 - \phi_0$ does not have a strong correlation with a_3/a or q .

Based on the behavior of the observed population density, we model the joint probability density of the multipole variables as

$$\begin{aligned} P(a_m/a, \phi_m - \phi_0, q) &= P(q) P(a_m/a, \phi_m - \phi_0 | q) \\ &= P(q) P(a_m/a | q) P(\phi_m - \phi_0 | a_m/a) \end{aligned} \quad (3.1)$$

by assuming that $\phi_m - \phi_0$ is conditionally independent of q , given a_m/a : $P(\phi_m - \phi_0 | a_m/a, q) = P(\phi_m - \phi_0 | a_m/a)$. There is no meaningful correlation between the two multipole pairs $(a_3/a, \phi_3 - \phi_0)$ and $(a_4/a, \phi_4 - \phi_0)$, so we assume conditional independence between them to build the probability density of the five variables:

$$\begin{aligned} P(a_3/a, \phi_3 - \phi_0, a_4/a, \phi_4 - \phi_0, q) &= P(q) P(a_3/a, \phi_3 - \phi_0 | q) P(a_4/a, \phi_4 - \phi_0 | q) \\ &= P(q) P(a_3/a | q) P(\phi_3 - \phi_0 | a_3/a) \\ &\quad P(a_4/a | q) P(\phi_4 - \phi_0 | a_4/a) \end{aligned} \quad (3.2)$$

The functional choices are:

- $P(q)$ — skew-normal distribution with three parameters: shape α_q , location ξ_q , and scale ω_q . Normalized within $(q_{\min} = 0.33, q_{\max} = 1)$ such that $\int_{q_{\min}}^{q_{\max}} P(q) dq = 1$.

⁴More exactly, H06 showed that a_4/a tends to have asymmetric distribution toward the positive, whereas β_4/a , α_3/a , and β_3/a do not. See Appendix A for the comparison of two multipole conventions, $(a_m/a, \phi_m - \phi_0)$ and $(\alpha_m/a, \beta_m/a)$.

⁵The original number of data points is 847, but 7 of them were disposed in conversion and error analysis; see Appendix C

⁶ $\theta_E \propto v_{\text{dis}}^2$ for a singular isothermal sphere, and thus the strong lensing area $\pi\theta_E^2 \propto v_{\text{dis}}^4$; refer to [56]

- $P(a_3/a | q)$ — Gaussian distribution with fixed mean $\mu = 0$ and standard deviation $\sigma(q)$, where $\sigma(q)$ is defined by linear interpolation through two control points. Normalized within $((a_3/a)_{\min} = -0.08, (a_3/a)_{\max} = 0.08)$.
- $P(\phi_3 - \phi_0 | a_3/a)$ — uniform distribution over the interval $[(\phi_3 - \phi_0)_{\min} = -\pi/6, (\phi_3 - \phi_0)_{\max} = \pi/6]$.
- $P(a_4/a | q)$ — skew-normal distribution with shape $\alpha(q)$, location $\xi(q)$, and scale $\omega(q)$, each specified by a three-point linear spline. Normalized within $((a_4/a)_{\min} = -0.05, (a_4/a)_{\max} = 0.14)$.
- $P(\phi_4 - \phi_0 | a_4/a)$ — symmetric generalized Gaussian distribution with fixed mean $\mu = 0$, and scale $\alpha(a_4/a)$ and shape $\beta(a_4/a)$, both described by three-point linear splines. Normalized within $[(\phi_4 - \phi_0)_{\min} = -\pi/8, (\phi_4 - \phi_0)_{\max} = \pi/8]$.

The limits $q_{\max} = 1$, $(\phi_m - \phi_0)_{\min} = -\pi/2m$, and $(\phi_m - \phi_0)_{\max} = \pi/2m$ are set by their mathematical definition, whereas q_{\min} , $(a_m/a)_{\min}$, $(a_m/a)_{\max}$, are set by the observed population limit; more specifically, maximum/minimum of data value ± 3 times of uncertainty.

The parameters were optimized by minimizing the sum of two Jensen–Shannon (JS) divergences between the model and the galaxy sample from [22] convolved with its uncertainties. The final JS divergence for $m = 3$ and $m = 4$ are $\text{JS}_{m=3} = 0.100$ and $\text{JS}_{m=4} = 0.143$, respectively. The details of optimization is shown in Appendix B. Table 1 lists the best-fit values that we use throughout this work.

Figure 3 compares the observed distributions in the [22] catalog (left panels), the best-fit model predictions (center), and their residuals (right) for both $m = 3$ and $m = 4$ multipoles in four axis-ratio slices. The general agreement between the observed and best-fit model

Table 1. Best-fit hyperparameters of the hierarchical prior. The total number of free parameters to be optimized are 37.

Group	Symbol	Value(s)
$P(q)$	α_q	-2.491
	ξ_q	0.910
	ω_q	0.175
$\sigma(q)$ spline ($m=3$)	$\sigma_{x,3}$	(0.482, 0.871)
	$\sigma_{y,3}$	(0.0075, 0.0043)
$\alpha(q)$ spline ($m=4$)	$\alpha_{x,4}$	(0.522, 0.614, 0.838)
	$\alpha_{y,4}$	(0.872, -1.054, -0.224)
$\xi(q)$ spline ($m=4$)	$\xi_{x,4}$	(0.437, 0.678, 0.851)
	$\xi_{y,4}$	(0.0485, 0.0124, 0.0008)
$\omega(q)$ spline ($m=4$)	$\omega_{x,4}$	(0.533, 0.628, 0.942)
	$\omega_{y,4}$	(0.0195, 0.0110, 0.0038)
$\alpha(a_4/a)$ spline	$\alpha'_{x,4}$	(0.0037, 0.0123, 0.0340)
	$\alpha'_{y,4}$	(0.540, 0.139, 0.068)
$\beta(a_4/a)$ spline	$\beta_{x,4}$	(0.0045, 0.0131, 0.0229)
	$\beta_{y,4}$	(1.929, 1.175, 1.601)

prediction demonstrates that the chosen functional forms are flexible enough to capture the empirical population while retaining analytic simplicity.

Algorithm 1 and Figure 4 demonstrate the sampling procedure. In our inference pipeline we do not draw these parameters directly; instead the joint prior of Eq. (3.2) is encoded as an additive log-likelihood evaluated for each MCMC proposal.

Algorithm 1 Draw a sample $(a_3/a, \phi_3 - \phi_0, a_4/a, \phi_4 - \phi_0, q)$. The same factorization is implemented as a custom prior-likelihood term to be fed into LENSTRONOMY and EMCEE pipeline. The sampling steps below show the conceptual sampling equivalent.

1. Sample $q \sim \text{SkewNormal}(\alpha_q, \xi_q, \omega_q)$.
 2.
 - a. Compute $\sigma(q)$ by linear interpolation of the two control points $(\sigma_{x,3}, \sigma_{y,3})$.
 - b. Sample $a_3/a \sim \mathcal{N}(0, \sigma(q)^2)$.
 - c. Sample $\phi_3 - \phi_0 \sim \mathcal{U}(-\pi/6, \pi/6)$.
 3.
 - a. Interpolate $(\alpha, \xi, \omega)(q)$ from their control points to obtain the shape of $P(a_4/a|q)$.
 - b. Sample $a_4/a \sim \text{SkewNormal}(\alpha(q), \xi(q), \omega(q))$.
 - c. Interpolate $\alpha(a_4/a)$ and $\beta(a_4/a)$ from their control points.
 - d. Sample $\phi_4 - \phi_0$ from the symmetric generalized Gaussian $G(\mu = 0, \alpha(a_4/a), \beta(a_4/a))$ on $[-\pi/8, \pi/8]$.
-

$$P(a_m/a, \phi_m - \phi_0, q \in q_{\text{range}})$$

q_{range} (0.33, 0.50] (0.50, 0.67] (0.67, 0.84] (0.84, 1.00]

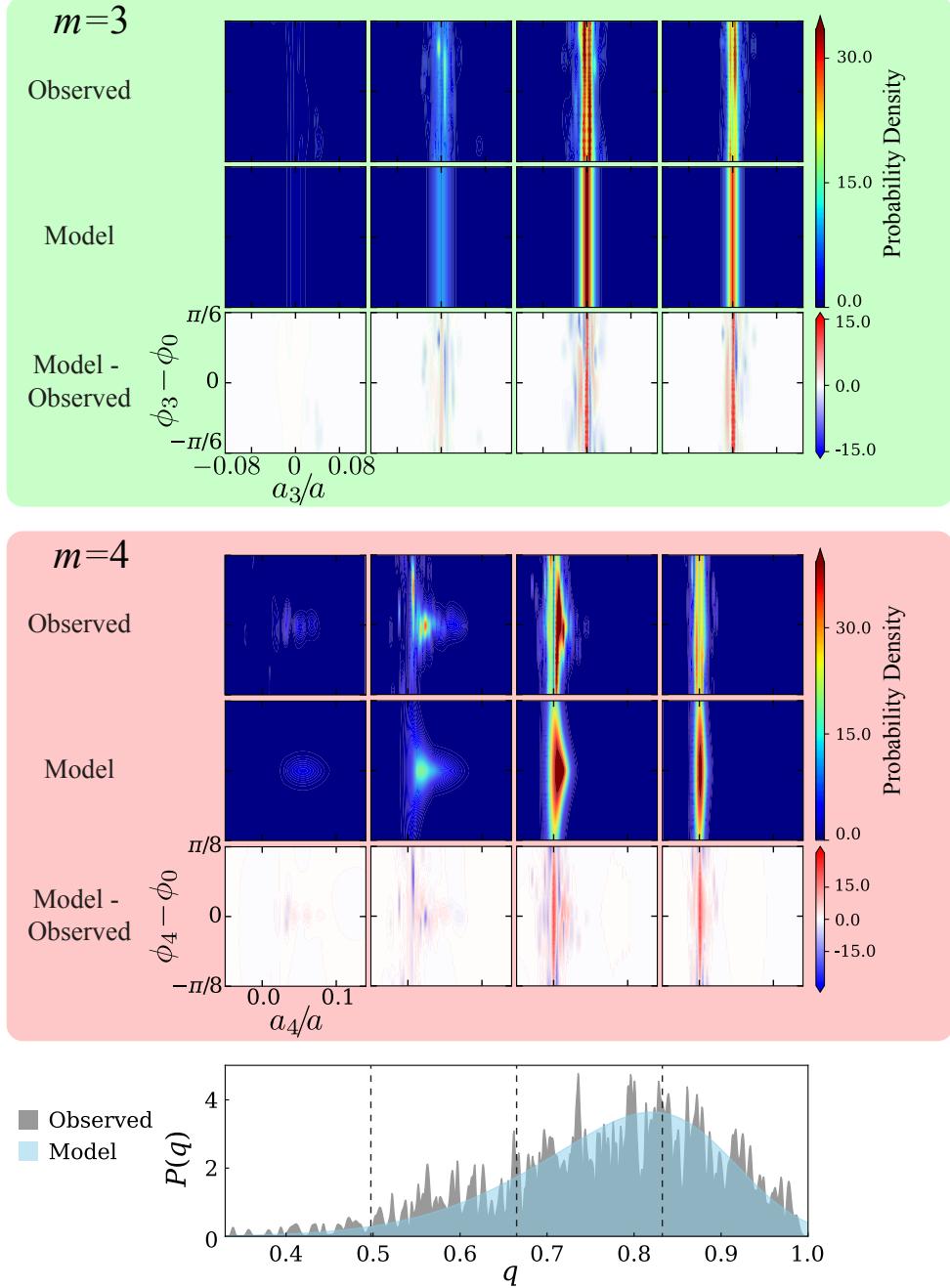
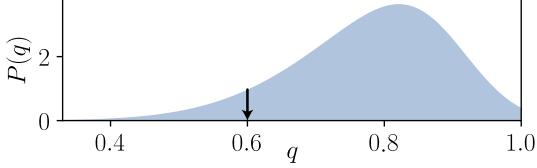


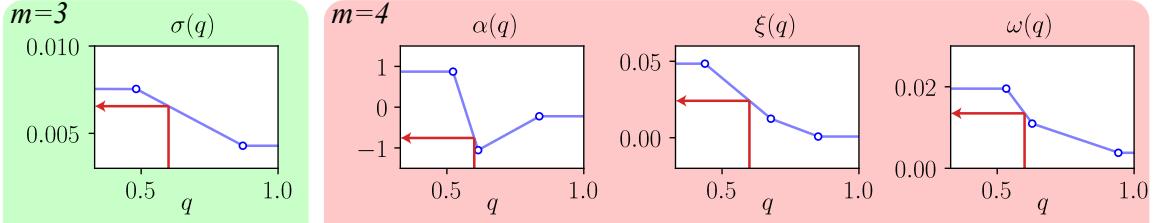
Figure 3. Observed versus modeled probability distributions for $(a_3/a, \phi_3 - \phi_0, q)$ (**top three rows**), $(a_4/a, \phi_4 - \phi_0, q)$ (**following three rows**), and $P(q)$ (**bottom**). Each triplet of heatmaps shows, from top to bottom, the observed kernel-density estimate (histogram of data points convolved with their uncertainties), the hierarchical model prediction using the best-fit parameters of Table 1, and their difference. Four columns correspond to non-overlapping slices in q ; they match with the four ranges shown in $P(q)$ with dashed line. Axis ticks are shared between heat maps and thus omitted from the panels to reduce visual clutter except one. Coordinate axes are consistent within each set of panels for $m = 3$ and $m = 4$.

1. Sample q from skew-normal $P(q)$

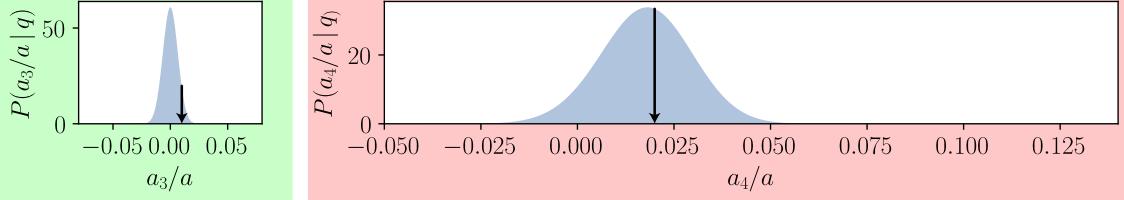


- Probability Density Functions
- ↓ Values Sampled from PDF
- Interpolation Control Points
- Interpolation Lines
- ← Interpolated Parameters

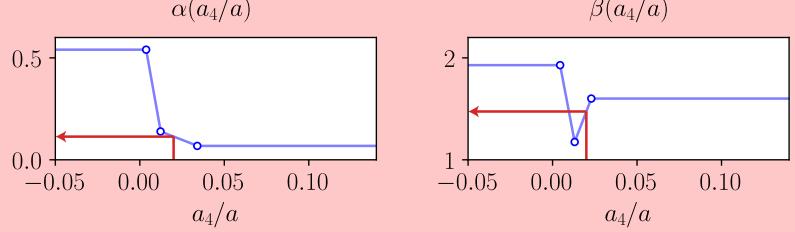
2. Interpolate parameters for $P(a_m/a|q)$ (σ for $m=3$; α, ξ, ω for $m=4$)



3. Sample a_m/a from $P(a_m/a|q)$ (Gaussian for $m=3$; skew-normal for $m=4$)



4. Interpolate parameters for $P(\phi_m - \phi_0 | a_m/a)$ (none for $m=3$; α, β for $m=4$)



5. Sample $\phi_m - \phi_0$ from $P(\phi_m - \phi_0 | a_m/a)$ (uniform for $m=3$; generalized Gaussian for $m=4$)

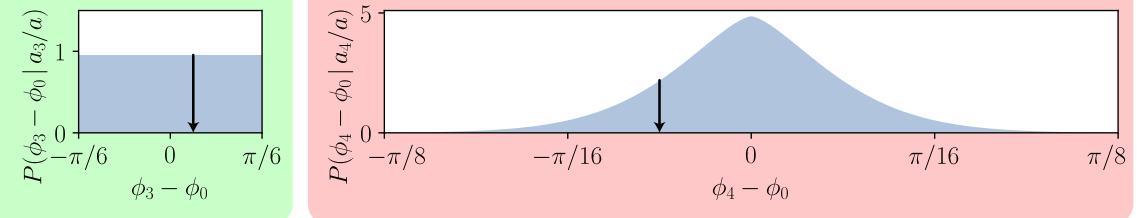


Figure 4. Graphical illustration of the hierarchical prior and the sampling steps summarized in Algorithm 1. Blue shaded curves show probability density functions; black arrows denote the sampled values; blue dots and lines are interpolation control points and interpolated values, respectively; red lines represent parameter interpolation.

4 Mock–data construction

Our goal is to provide a controlled, HST-quality reference case in which the impact of multipole priors and extended–arc information can be quantified unambiguously. Therefore, we fix the macro model parameters and source morphology based on a single, well-studied system with clear extended arcs: WGD 2038-4008 [16, 57]. This system has a source redshift of $z_{\text{source}} = 0.777$, $z_{\text{lens}} = 0.230$ [57, 58]. We fitted the original HST F814W image of this system with a model without multipoles—EPL+shear for lens mass, Sérsic for lens light and source light profiles, and point-like quasars—to set the realistic model parameters. This ensures that in addition to realistic signal to noise, our mock lenses will have typical properties of gravitational lenses including potential contamination from the deflector light and realistic extended quasar host galaxy brightness. We add realistic mass multipole parameters on the fitted model based on two multipole scenarios. Investigation for multiple mock/real lens systems is left to future work. Below we summarize the ingredients.

4.1 Parameter setting

Multipole scenarios We base our parameter choices within a reasonable prior shown in Section 3. We test two scenarios for the multipoles of the main deflector as described in Table 2. One scenario, ‘Mild $m = 4$ ’ has $a_4/a = 0.01$. The other scenario, ‘Strong $m = 4$ ’ has $a_4/a = 0.03$. In both cases, we align the $m = 4$ multipole with EPL; $\phi_4 - \phi_0 = 0$ (rad), while keeping the prior broad. For $m = 3$ multipoles, we set them to be $a_3/a = 0.01$ and $\phi_3 - \phi_0 = 0.2$ (rad) for both cases.

These multipole scenarios are selected as they are representative of realistic combinations of parameter values seen in galaxy isophotes, and they span a possible range of perturbations to the lens models.

Observation scenarios For each multipole scenario described above, we made two simulated observations of the lens system; one in which only point sources were imaged, and the other with point sources and extended arcs. The mock observation data was generated using LENSTRONOMY. Figure 5 gives an example of two of our mock data sets.

Other parameters The parameter values and priors for other mass and light components are summarized in Tables 5 and 6. Note that, for the *point-sources only* data sets the host light is not included.

Multipole Scenario	Parameter Name	Simulation Truth	Fitting Prior
Mild $m = 4$	a_3/a	0.01	
	$\phi_3 - \phi_0$	0.2 (rad)	Jointly sampled
	a_4/a	0.01	(see Section 3)
	$\phi_4 - \phi_0$	0 (rad)	
Strong $m = 4$	a_3/a	0.01	
	$\phi_3 - \phi_0$	0.2 (rad)	Jointly sampled
	a_4/a	0.03	(see Section 3)
	$\phi_4 - \phi_0$	0 (rad)	

Table 2. $m=3, 4$ multipole parameters with two different scenarios.

4.2 Data quality

We create mock observations based on Hubble Space Telescope’s WFC3 imaging data from programs GO-15320 and GO-15652 (PI: Treu, T.). Those programs carried out a uniform multi-band imaging campaign of 31 quadruply imaged quasars and [16] provided the optimal parameter for modeling our mock lens system, WGD 2038-4008. We base our simulations on the observations with F814W which had the best combination of sensitivity and clear extended arcs.

The exposure time is set to be 920 seconds, the pixel size is set to be 0.04 arcseconds, and the background noise level is set to be 0.006 photons/second for each pixel. The point spread function (PSF) is modeled as a two-dimensional Gaussian with a known full width at half maximum (FWHM) of $0''.10$. This choice is a simplification; in real observations, the PSF is typically more complex and must be inferred from the data. We opt not to model a complex PSF or reconstruct it from the data to avoid PSF-related complications. Section 5 now treats these mocks as data and applies the modeling framework of Section 2.

5 Inference framework

5.1 Likelihood and priors

We infer the lens model parameters from the mock data following the standard LENSTRONOMY fitting procedure [38, 39] together with two custom prior likelihoods. The total log-likelihood

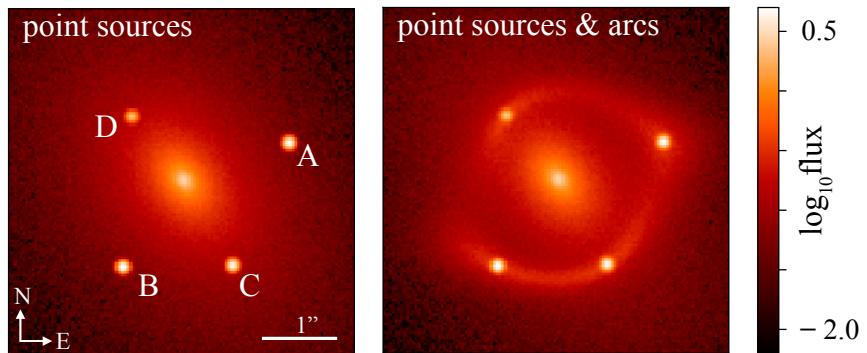


Figure 5. Comparison of simulations of HST-quality quadruply imaged quasars with point sources only (**left**) and point sources and extended arcs (**right**). The four quasar images are marked as A, B, C, and D. Note that the ‘point sources only’ case uses limited information (point source positions) only, whereas the ‘point sources and extended arcs’ case corresponds to an analysis using complete information with extended arcs. This way we can measure how much improvement comes when the full observation with extended arcs is modeled, compared to using images positions only. The change from one multipole scenario to another is quite subtle (e.g. slight shift of point sources and arcs), and thus not compared here.

of the inference process is given as

$$\begin{aligned} \ln \mathcal{L}_{\text{tot}} = & \ln \mathcal{L}_{\text{img}} + \ln \mathcal{L}_{\text{ps}} + \underbrace{\ln \mathcal{L}_e}_{\text{ellipticity prior, Appendix E}} \\ & + \underbrace{\ln P_{\text{model}}(a_3/a, \phi_3 - \phi_0, a_4/a, \phi_4 - \phi_0, q)}_{\text{population prior, Section 3}} \end{aligned} \quad (5.1)$$

where:

- \mathcal{L}_{img} is the pixel-based image likelihood. Note that the point-source *fluxes* are *not* compared to the observations as done by other studies of lensed quasars using HST [e.g. 18, 57]; quasar image flux in HST optical broad-band imaging is dominated by emission from the quasar accretion disk and thus sensitive to stellar microlensing in reality.
- \mathcal{L}_{ps} quantifies how well the multiple lensed quasar images can be mapped back to a common point in the source plane under the current lens model.
- \mathcal{L}_e is an ellipticity prior that makes uniform prior on (q, ϕ_0) . See Appendix E for details.
- P_{model} is the joint prior for multipole parameters and q described in Section 3.

Other parameters have broad uniform priors (see Tables 5 and 6). The true PSF is assumed to be known, and thus not optimized.

5.2 Sampling strategy

The posterior landscape is explored in two stages:

- (i) **Particle–swarm optimization** (PSO), 100 particles, 200 iterations, identifies a high-likelihood region.
- (ii) **Ensemble MCMC** (the `emcee` sampler) with a walker-to-parameter ratio of 8:1, 40 000 steps each. The first 30 % of the chain is discarded as burn-in; out the remaining chain, $\sim 500\,000$ samples are used for all posterior plots and flux–ratio statistics.

5.3 Derived quantities

For each retained sample we evaluate the signed magnifications

$$\mu(\vec{\theta}) = \frac{1}{(1 - \kappa)^2 - (\gamma_1^2 + \gamma_2^2)}, \quad (5.2)$$

where $\kappa = \frac{1}{2} \left[\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} \right]$, $\gamma_1 = \frac{1}{2} \left[\frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi}{\partial y^2} \right]$, and $\gamma_2 = \frac{\partial^2 \psi}{\partial x \partial y}$ for the projected gravitational potential ψ . The three model-predicted flux ratios were calculated as $B/A = |\mu_B/\mu_A|$, $C/A = |\mu_C/\mu_A|$, $D/A = |\mu_D/\mu_A|$.

For each multipole scenario, we define the with-arcs precision improvement factor, $\mathcal{F}_{(p)}$, as the ratio between the 68% confidence interval uncertainties of a variable p for ‘point source only’ case, $\Delta p|_{\text{point sources}}$, and that of ‘point sources and arcs’ case, $\Delta p|_{\text{point sources \& arcs}}$.

$$\mathcal{F}_p \equiv \frac{\Delta p|_{\text{point sources}}}{\Delta p|_{\text{point sources \& arcs}}}. \quad (5.3)$$

6 Results

The posterior comparison of *point source only* (red) and *point sources + arcs* (blue) are shown in Figures 6 and 7 for the multipole parameters and in Figure 8 for the flux ratios. The improvement in parameter precision is summarized in Tables 3 and 4.

The axis ratio q is significantly better constrained when extended arcs are included, with the precision improvement factor \mathcal{F}_q exceeding 10 in both multipole scenarios. This more narrowly defined q distribution in turn enables sharper joint priors on the multipole parameters.

The amplitudes of the multipoles, a_3/a and a_4/a , are largely unconstrained by point source positions alone; their posteriors remain broad and prior-dominated. When arcs are added, however, these parameters become well-constrained, with precision improvement factors of $\mathcal{F}_{a_3/a} \approx \mathcal{F}_{a_4/a} \approx 3.5$.

The multipole angles $\phi_3 - \phi_0$ and $\phi_4 - \phi_0$ also show clear improvement in constraint when arcs are included. In the mild $m = 4$ case, the inclusion of arcs improves the constraint on $\phi_4 - \phi_0$ by factors of $\mathcal{F}_{\phi_3 - \phi_0} \approx \mathcal{F}_{\phi_4 - \phi_0} \approx 7.5$, and in the strong $m = 4$ case, the factors increase to ≈ 12 . This indicates that arc information becomes especially powerful when the underlying multipole amplitude is more pronounced.

Lastly, the inclusion of arcs also greatly sharpens the predicted flux ratios. On average, the uncertainties on B/A , C/A , and D/A are reduced by a factor of approximately 6, regardless of the underlying multipole scenario. This substantial improvement in predicted flux precision underscores the benefit of arc information for studies relying on flux ratio anomalies.

In summary, when multipole amplitudes are at the typical level observed in the isophotes of massive elliptical galaxies, extended arcs are essential for deriving accurate and precise constraints on the smooth lens model. Point sources alone are insufficient for this task, but their combination with arcs enables strong recovery of both multipole parameters and flux ratios. This improvement comes in part from constraining the axis ratio q .

7 Discussion and conclusion

Extended arcs are essential to constrain multipole parameters. Adding the host-galaxy arcs improves the precision of the model flux ratios by an order of magnitude and reveals the multipole amplitude and orientation.

This work provides a reasonable prior on multipole parameters for the macro model, which can increase the robustness of strong lens modeling by incorporating well-motivated complexity. Future strong lens analyses—including time-delay measurements and substructure studies—can build on this framework. In particular, for studies that use flux ratio anomalies to infer substructures [12–14, 59], our results suggest that incorporating extended arcs can improve both the precision and robustness of the analysis, while giving meaningful constraints on multipole parameters.

We adopted the circular multipole convention to remain consistent with H06, but we regard the elliptical convention as the preferred choice for future work because it better reflects realistic galaxy shapes as discussed in Section 2.1.2 [41, 42]. Circular and elliptical multipoles coincide when $q = 1$, and deviate only mildly when either $q \sim 1$ or the multipole amplitudes are small ($|a_m/a| \ll 1$). In our sample, the majority of systems satisfy $|a_m/a| < 0.02$ (98.5% for $m = 3$, 87.6% for $m = 4$, 86.7% together), while noticeably high values of a_4/a appear in

Multipole Parameters	a_3/a	$\phi_3 - \phi_0$	a_4/a	$\phi_4 - \phi_0$	q
Mild $m = 4$					
Truth	0.010	0.200	0.010	0.000	0.580
Point sources	$0.001^{+0.005}_{-0.005}$	$0.036^{+0.316}_{-0.382}$	$0.003^{+0.011}_{-0.007}$	$0.001^{+0.214}_{-0.210}$	$0.750^{+0.086}_{-0.090}$
Point sources & arcs	$0.009^{+0.001}_{-0.001}$	$0.214^{+0.050}_{-0.042}$	$0.012^{+0.002}_{-0.002}$	$-0.034^{+0.023}_{-0.033}$	$0.588^{+0.008}_{-0.007}$
With-arcs precision improvement factor	3.6	7.5	4.6	7.6	11.6
Strong $m = 4$					
Truth	0.010	0.200	0.030	0.000	0.580
Point sources	$0.000^{+0.005}_{-0.005}$	$0.027^{+0.322}_{-0.372}$	$0.002^{+0.008}_{-0.006}$	$-0.017^{+0.254}_{-0.228}$	$0.786^{+0.101}_{-0.111}$
Point sources & arcs	$0.013^{+0.001}_{-0.001}$	$0.176^{+0.033}_{-0.029}$	$0.030^{+0.002}_{-0.002}$	$-0.022^{+0.020}_{-0.019}$	$0.584^{+0.009}_{-0.008}$
With-arcs precision improvement factor	3.1	11.2	3.5	12.6	12.2

Table 3. Comparison of truth values, inferred parameters, and precision improvements of multipole parameters for the “Mild $m = 4$ ” and “Strong $m = 4$ ” scenarios. Each row shows the true value, the inferred median and 68% confidence intervals for the point sources only and point sources & arcs cases, and the resulting precision improvement factor from adding arcs. This factor is defined as the ratio of the confidence interval width in the point-source-only case to that in the combined case.

Flux Ratios	B/A	C/A	D/A
Mild $m = 4$			
Truth	1.234	0.828	0.419
Point sources	$1.199^{+0.060}_{-0.054}$	$1.061^{+0.100}_{-0.150}$	$0.488^{+0.046}_{-0.053}$
Point sources & arcs	$1.226^{+0.014}_{-0.014}$	$0.826^{+0.013}_{-0.013}$	$0.417^{+0.007}_{-0.006}$
With-arcs precision improvement factor	4.1	9.5	7.7
Average improvement factor: 6.7			
Strong $m = 4$			
Truth	1.235	0.751	0.389
Point sources	$1.192^{+0.055}_{-0.057}$	$1.118^{+0.090}_{-0.138}$	$0.478^{+0.037}_{-0.048}$
Point sources & arcs	$1.225^{+0.019}_{-0.018}$	$0.745^{+0.013}_{-0.012}$	$0.397^{+0.006}_{-0.006}$
With-arcs precision improvement factor	2.997	9.045	6.817
Average improvement factor: 5.7			

Table 4. Same as Table 3 but of flux ratios. The average improvement factor (gray cells) summarizes the relative gain in precision for the flux ratio predictions.

'Mild $m = 4$ ' Inference Result

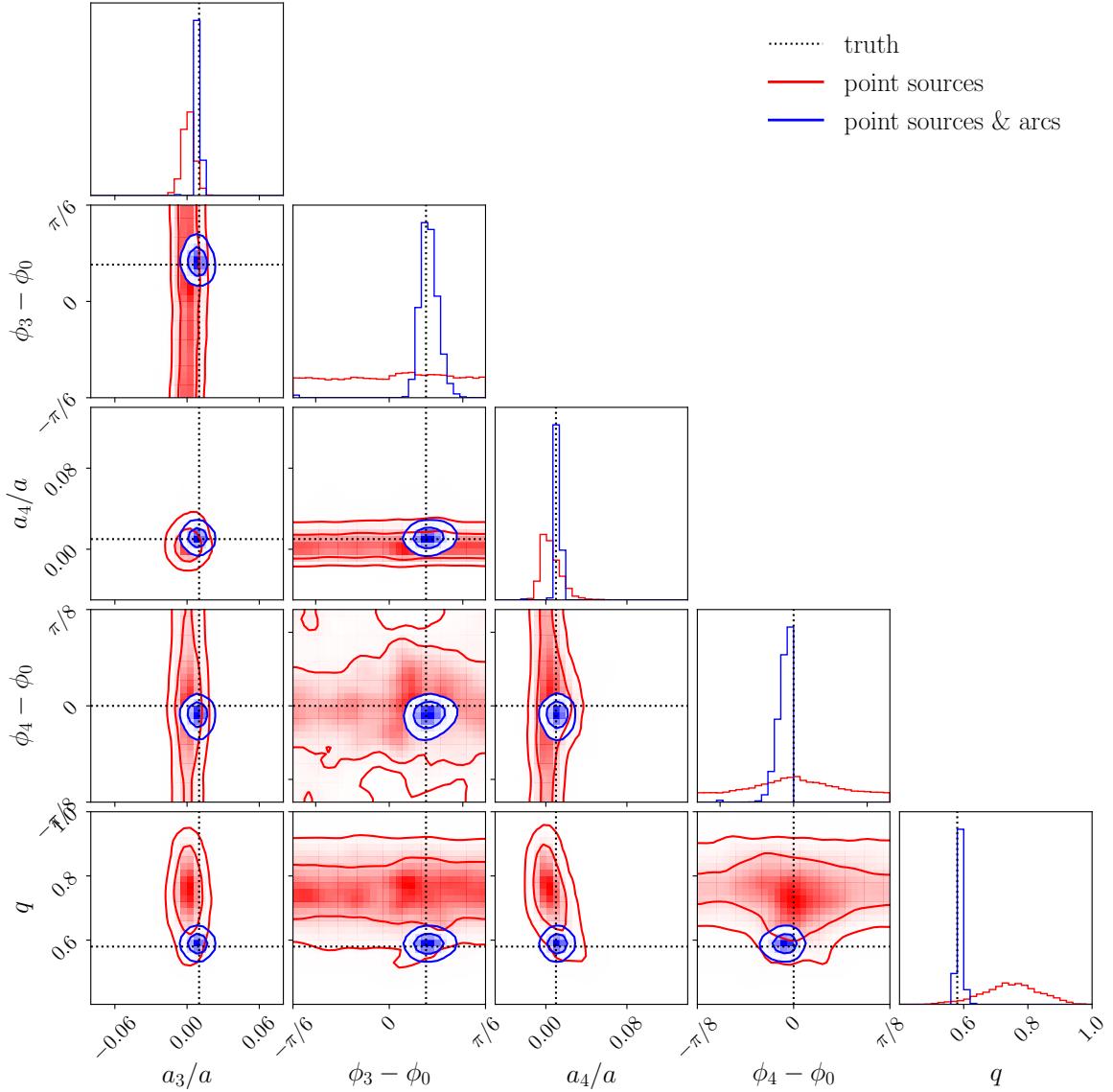


Figure 6. Corner plot of the joint posterior for the multipole parameters in the 'Mild $m = 4$ ' mock lens. Red contours/histograms correspond to the point-source-only fit, while blue show the fit that also uses the extended arcs. Shaded regions (light to dark) enclose the 68% and 95% highest-posterior-density levels. Black dashed lines indicate the truth values. Numerical summaries are given in Table 3.

'Strong $m = 4$ ' Inference Result

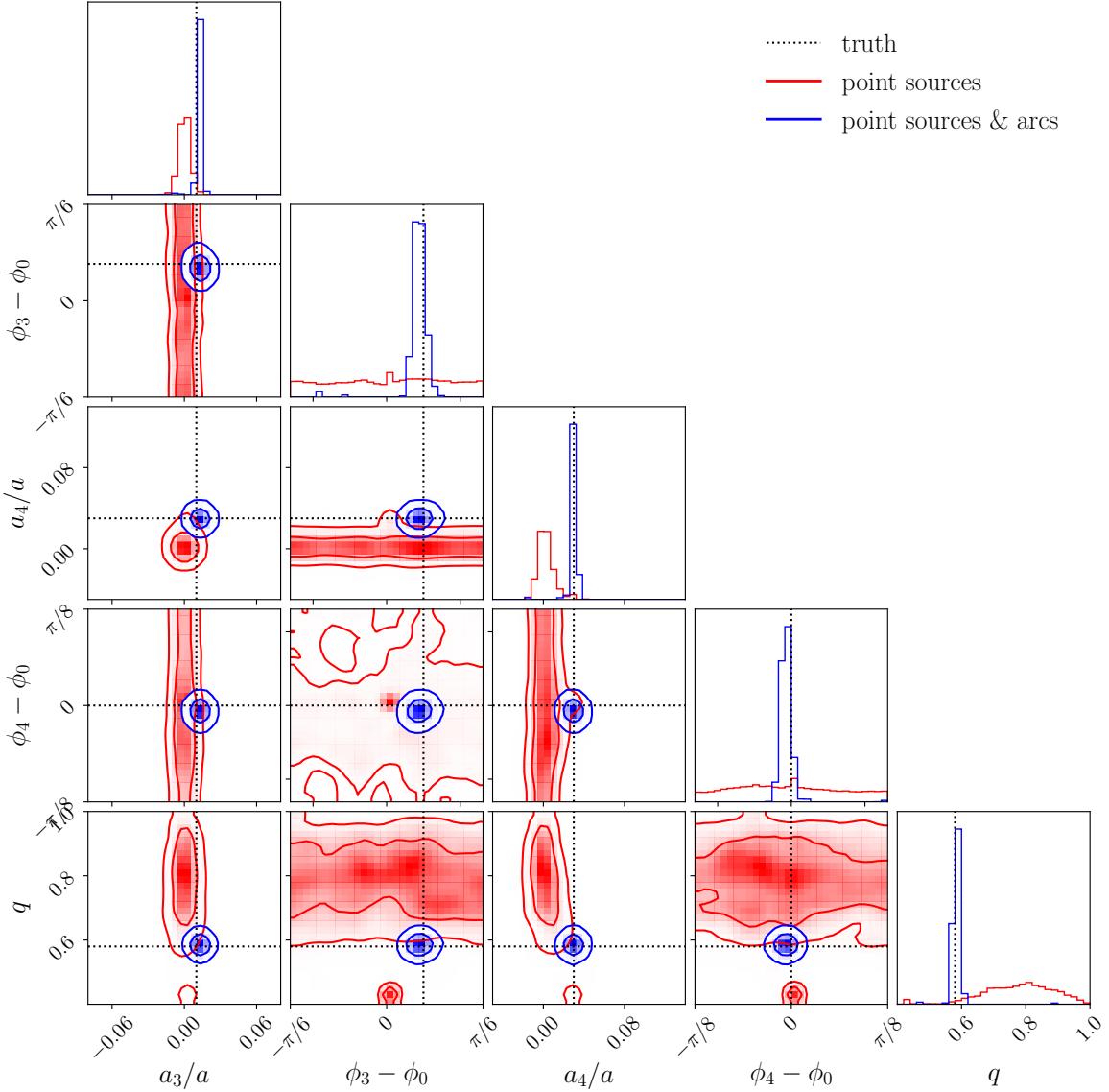


Figure 7. Same as Figure 6 but for ‘Strong $m = 4$ ’ mock lens.

highly elliptical lenses. Consequently, the circular multipole prior presented here may serve as a temporary substitute for an elliptical multipole prior, particularly systems with $q \sim 1$, until future optical surveys measured in the elliptical multipole convention enable the construction of a dedicated prior.

An additional implication is about the issue of shear bias. It has been shown that the external shear fitted to strong lensing systems often compensates for missing $m = 4$ multipole terms rather than representing true cosmic shear [29]⁷. Including $m = 4$ multipoles in the lens model has been proposed as one solution to the lack of correlation between external shear

⁷The truncation of the lens mass map has also been noted as a factor influencing shear measurements; see [60].

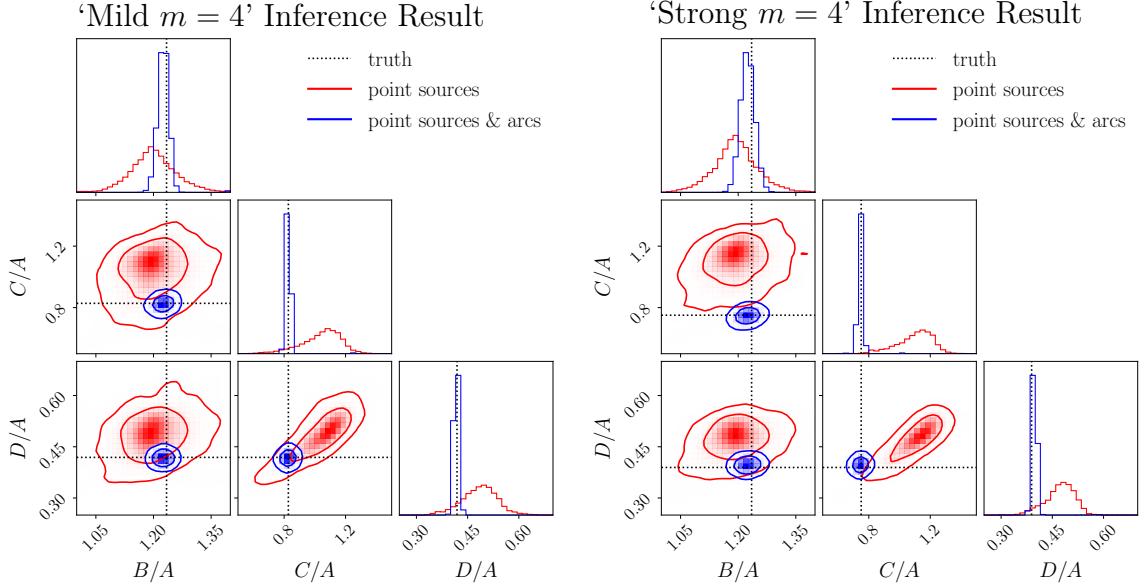


Figure 8. Same as Figures 6 and 7 but for flux ratios. Numerical summaries are given in Table 4.

measured in strong lens systems and cosmic shear measured in weak lensing. Our findings imply that multipole terms can be accurately constrained when arcs and a joint prior on multipoles are included. This approach could help eliminate the shear bias in future studies.

There are, however, limitations and opportunities for further development. This proof-of-concept study is based on a single lens geometry and assumes a Gaussian PSF. Several future directions are clear: (i) applying the method to multiple mock lenses with different image configurations and source morphologies, (ii) applying it to quadruply lensed quasars with arcs observed with HST, (iii) incorporating a more complex source light model, and (iv) combining the method with JWST mid-IR flux ratios in a joint macro + substructure analysis. It is important to note that our current analysis has several simplifications and does not include dark matter substructures in the mock observations or in the modeling. Therefore, although we observe significant improvement in constraining the lens model using imaging data and arcs, this does not guarantee the same level of improvement when dark matter substructures are present.

A companion paper [20] extends this work by incorporating full realizations of dark matter substructures and modeling multipoles of $m = 3, 4$ within a comprehensive lensing framework, while the joint multipole prior proposed in this study was not used. Their inference is performed directly on imaging arcs using shapelet-based source reconstructions with substructure lensing together. However, the use of highly flexible source models and inclusion of dark matter substructure introduces significant degeneracies between lens and source parameters. As a result, the posterior constraints on the multipole parameters are noticeably broader. This suggests that flexible source and inclusion of dark matter substructures can dilute the constraining power of arcs on the lens potential unless additional priors or constraints are imposed. In contrast, our controlled mock setup—with a simple source and clean arc morphology—highlights the maximum potential gain from arcs in constraining multipoles under idealized conditions.

In summary, in this work, the population prior for $(a_3/a, \phi_3 - \phi_0, a_4/a, \phi_4 - \phi_0, q)$ based

on optical observations is designed and implemented into the strong lens system. When combined with extended arcs, yields an order-of-magnitude improvement in the precision of smooth-model multipole parameters and flux ratios. This approach may also contribute to future strong lensing studies and enable separation of line-of-sight (cosmic) external shear from apparent “shear” that is actually produced by higher-order multipoles in the deflector.

Acknowledgments

This research was conducted using MERCED cluster (NSF-MRI, #1429783) and Pinnacles (NSF MRI, #2019144) at the Cyberinfrastructure and Research Technologies (CIRT) at University of California, Merced. We thank Caina Hao and Shude Mao for providing the original data from their optical survey of galaxy multipoles, Anowar Shajib for providing model parameters for WGD2038-4008, and Lyne Van de Vyvere for a discussion about multipole notation conventions. AMN and MO acknowledge support by the NSF through grant 2206315 "Collaborative Research: Measuring the physical properties of dark matter with strong gravitational lensing", and from #GO-2046 which was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127. DG acknowledges support from a Brinson Prize Fellowship. SB acknowledges support from the Department of Physics & Astronomy, Stony Brook University.

A Comparison of multipole conventions

In this paper, the multipole radial deviation of the isophotal or isodensity contour from the best-fit ellipse was expressed as a single cosine function with a multipole phase ϕ_m as follows.

$$\delta r = a_m \cos(m(\phi - \phi_m)) \quad (\text{A.1})$$

$\cdots (a_m, \phi_m)$ convention

Here, r and ϕ are the typical radial and angular coordinates, $r = \sqrt{x^2 + y^2}$ and $\phi = \arctan 2(y, x)$. The same equations can be converted into a different convention following [22, 61] using the sum of cosine and sine functions as follows⁸.

$$\delta r = \alpha_m \cos(m(\phi - \phi_0)) + \beta_m \sin(m(\phi - \phi_0)) \quad (\text{A.2})$$

$\cdots (\alpha_m, \beta_m)$ convention

Note that when the multipole and the ellipse are aligned, $\phi_m - \phi_0 = 0$ in the first convention, $\beta_m = 0$ in the second convention, and $a_m = \alpha_m$.

The conversion from (a_m, ϕ_m) to (α_m, β_m) is given as follows, from the angle sum formula for cosine.

$$\alpha_m = a_m \cos(m(\phi_m - \phi_0)), \quad \beta_m = a_m \sin(m(\phi_m - \phi_0)) \quad (\text{A.3})$$

$\cdots (a_m, \phi_m) \rightarrow (\alpha_m, \beta_m)$

The other way of conversion from (α_m, β_m) to (a_m, ϕ_m) is not unique. We choose to do it by the following.

$$a_m = \text{sign}(\alpha_m) \sqrt{\alpha_m^2 + \beta_m^2}, \quad \phi_m = \phi_0 + \frac{1}{m} \arctan(\beta_m / \alpha_m) \quad (\text{A.4})$$

$\cdots (\alpha_m, \beta_m) \rightarrow (a_m, \phi_m)$

This way of conversion lets a_m keeps the sign of α_m and its significance; e.g. $a_4 > 0$ means disk-like and $a_4 < 0$ means boxy. If a different conversion rule is used, this property is not guaranteed. For example, assume the following conversion: $a'_m = \sqrt{\alpha_m^2 + \beta_m^2}$ and $\phi'_m = \phi_0 + \frac{1}{m} \arctan 2(\beta_m, \alpha_m)$. In this case, a'_m is always non-negative and the ‘boxy/diskyness’ of $m=4$ multipole depends on the range of the misalignment $\phi_4 - \phi_0$, which is more tricky to recognize.

B Optimization Details

The loss function to be minimized in the optimization process is defined as the sum of two JS divergences, one for $m = 3$ and the other for $m = 4$:

⁸The equations in [22] do not have the angle of the ellipse ϕ_0 because their coordinate system is aligned with the elliptical profile; i.e. $\phi_0 = 0$ by construction. Here we included it for generality of the equation. Also, they use (a_m, b_m) instead of (α_m, β_m) ; we are using the latter to distinguish the conventions.

$$\begin{aligned}
L_{\text{JS}} &= \text{JS}_{m=3} + \text{JS}_{m=4} \\
&= \text{JS}[P_{\text{obs}}(a_3/a, \phi_3 - \phi_0, q) \| P_{\text{mod}}(a_3/a, \phi_3 - \phi_0, q)] + \\
&\quad \text{JS}[P_{\text{obs}}(a_4/a, \phi_4 - \phi_0, q) \| P_{\text{mod}}(a_4/a, \phi_4 - \phi_0, q)].
\end{aligned} \tag{B.1}$$

The optimization process optimized all variables shown in Table 1 using the ADAM optimizer in PyTorch, where P_{obs} is a kernel-density estimate of the sample from H06 [22]. The model likelihood P_{mod} is evaluated by Eq. (3.2). Each galaxy is represented by a 3-D Gaussian centered at the measured value, with widths given by its quoted uncertainties. For angular variables ($\phi_3 - \phi_0$, $\phi_4 - \phi_0$), we replicate the kernel across the periodic boundaries⁹.

JS divergence is defined as

$$\text{JS}[P \| Q] = \frac{1}{2} \text{KL}[P \| M] + \frac{1}{2} \text{KL}[Q \| M], \quad \text{where } M = \frac{1}{2}(P + Q), \tag{B.2}$$

and the Kullback–Leibler divergence is given by

$$\text{KL}(P \| Q) = \int P(x) \log \frac{P(x)}{Q(x)} dx_1 dx_2 \cdots dx_n. \tag{B.3}$$

The JS divergence provides a symmetric and smooth measure of dissimilarity between probability distributions and is widely used in the optimization of probabilistic models. JS divergence has the lower bound of 0 when $P(x)$ and $Q(x)$ are identical and the upper bound of $\log(2) \approx 0.69$ when $P(x)$ and $Q(x)$ are completely disjoint. After optimization procedure¹⁰, it reaches to $\text{JS}_{m=3} = 0.11$ and $\text{JS}_{m=4} = 0.16$.

C Generation of the observational probability densities

The [22] catalog provides (α_m, β_m, q) for $m = 3$ and $m = 4$ together with 1σ measurement uncertainties $(\Delta\alpha_m, \Delta\beta_m, \Delta q)$. This is converted to $(a_m/a, \phi_m - \phi_0, q)$ using the conversion shown in Appendix A. The uncertainties $(\Delta a_m/a, \Delta\phi_m - \phi_0, \Delta q)$ are calculated using linear error analysis, but there were data points showing imaginary uncertainties from linear analysis and their uncertainties were calculated again quadratically. There were 7 out of 847 data points that still showed imaginary uncertainties even then, which were removed from this analysis.

To compare our hierarchical model with the data, we convert this discrete sample into two continuous and normalized probability densities $P_{\text{obs}}(a_3/a, \phi_3 - \phi_0, q)$ and $P_{\text{obs}}(a_4/a, \phi_4 - \phi_0, q)$. Our procedure is summarized below.

1. **Gaussian kernel.** Each galaxy i is represented by a 3-D Gaussian

$$G_i = \mathcal{N}((a_m/a)_i, \Delta(a_m/a)_i^2) \mathcal{N}((\phi_m - \phi_0)_i, \Delta(\phi_m - \phi_0)_i^2) \mathcal{N}(q_i, \Delta q_i^2).$$

2. **Angular wrapping.** Because the angular variables are periodic, $\phi_m - \phi_0 \in (-\frac{\pi}{2m}, \frac{\pi}{2m})$, a kernel whose width extends beyond the boundary must be replicated. We compute $n = \lceil |(\phi_m - \phi_0)_i \pm 3\Delta(\phi_m - \phi_0)_i| / (\pi/2m) \rceil$ and create shifted copies $((a_m/a)_i', (\phi_m - \phi_0)_i', q_i') = (\pm(a_m/a)_i, (\phi_m - \phi_0)_i \pm 2k \cdot \pi/2m, q_i)$ for every $k = -n, \dots, n$. Odd- k shifts flip the sign of a_m/a because the m -th multipole is antisymmetric under a half-period rotation.

⁹For $\phi_3 - \phi_0$ the fundamental domain is $(-\pi/6, \pi/6)$, for $\phi_4 - \phi_0$ it is $(-\pi/8, \pi/8)$. Any kernel that crosses a boundary is mirrored with the appropriate parity in a_3/a or a_4/a ; see Appendix C.

¹⁰5000 epochs with adaptive learning rate (ReduceLROnPlateau) starting with 0.01 with patience 50 and factor 0.5

3. **Weighting.** Each replica is assigned the analytical weight $w_k = \text{CDF}_{\mathcal{N}}(\frac{\pi}{2m} + 2k\frac{\pi}{2m}) - \text{CDF}_{\mathcal{N}}(-\frac{\pi}{2m} + 2k\frac{\pi}{2m})$. This ensures that the integral of the replicated kernels equals unity. All kernels share the same weight in $\phi_m - \phi_0$ and q .
4. **Normalization.** The sum $P_{\text{obs}} = \sum_i \sum_k w_{ik} G_{ik}$ is evaluated on a regular $(100)^3$ grid and renormalised so that $\int P_{\text{obs}} = 1$.

D Modeling details

Table 5 and 6 provide the lens and light parameters other than the multipole parameters used to create the mock data.

Profile	Parameter Name	True Value	Prior	Note
Elliptical Power Law (EPL)	θ_E	1''.37	$U(0, 10)$	
	γ	2.50	$U(1.5, 2.5)$	
	x_{center}	0''.043	$U(-10, 10)$	
	y_{center}	0''.002	$U(-10, 10)$	
	e_1	-0.07		See Appendix E
	e_2	-0.25		
	q	0.58	Converted from (e_1, e_2)	(a)
External Shear	ϕ_0	-0.93 (-53°)	Converted from (e_1, e_2)	(b)
	γ_1	0.04	$U(-0.5, 0.5)$	
	γ_2	0.10	$U(-0.5, 0.5)$	
	γ_{ext}	0.10	Converted from (γ_1, γ_2)	(c)
$m = 3$ Elliptical Multipole	ϕ_{ext}	0.60 (34°)	Converted from (γ_1, γ_2)	(d)
	a_3/a		See Table 2	
	$\phi_3 - \phi_0$			
	x_{center}	0''.043	Jointly sampled with	
$m = 4$ Elliptical Multipole	y_{center}	0''.002	EPL's $(x_{\text{center}}, y_{\text{center}})$	
	a_4/a		See Table 2	
	ϕ_4			
	x_{center}	0''.043	Jointly sampled with	
	y_{center}	0''.002	EPL's $(x_{\text{center}}, y_{\text{center}})$	

Table 5. The true values and priors of the lens mass parameters used for simulation and fitting of the lensed quasar system.

(a) $q = \frac{1-c}{1+c}$, $c = \sqrt{e_1^2 + e_2^2}$. (b) $\phi_0 = \frac{1}{2} \arctan 2(e_2, e_1)$. (c) $\gamma_{\text{ext}} = \sqrt{\gamma_1^2 + \gamma_2^2}$. (d) $\phi_{\text{ext}} = \frac{1}{2} \arctan 2(\gamma_2, \gamma_1)$.

E Ellipticity parameterization and prior

The multipole prior $P(a_3/a, \phi_3 - \phi_0, a_4/a, \phi_4 - \phi_0, q)$ discussed in Section 3 provides a skew-normal distribution as the prior for q , and this is implemented for the prior of (e_1, e_2) , considering their relationship with q , which is $q = \frac{1-c}{1+c}$, $c = \sqrt{e_1^2 + e_2^2}$. However, this is not enough because the parameterization of (e_1, e_2) itself carries a strong bias toward small q and thus affect the multipole prior too, if not adjusted. In the following content, we explain this problem and an additional prior to relieve this bias.

Kind	Parameter Name	True Value	Prior	Note
Quasar	x_{source}	0''.18	Not directly sampled	(a)
	y_{source}	-0''.10	Not directly sampled	
Elliptical Sérsic (Source Light, when arcs exist)	I_e	40	Not directly sampled	
	R_{sersic}	0''.37	$U(0.001, 10)$	
	n_{sersic}	1.0	$U(0.5, 5)$	
	x_{source}	0''.18	Jointly sampled with	
	y_{source}	-0''.10	Quasar's ($x_{\text{source}}, y_{\text{source}}$)	
	e_1	0.37	$U(-0.5, 0.5)$	
Elliptical Sérsic (Lens Light)	e_2	0.13	$U(-0.5, 0.5)$	
	q_{source}	0.43	Converted from (e_1, e_2)	(b)
	ϕ_{source}	0.17 (9.7°)	Converted from (e_1, e_2)	(c)
	I_e	12	Not directly sampled	
	R_{sersic}	3.3	$U(0.001, 10)$	
	n_{sersic}	3.9	$U(0.5, 5)$	
	x_{lens}	0''.03	$U(-10, 10)$	
	y_{lens}	0''.01	$U(-10, 10)$	
	e_1	-0.05	$U(-0.5, 0.5)$	
	e_2	-0.18	$U(-0.5, 0.5)$	
	q_{lens}	0.69	Converted from (e_1, e_2)	(b)
	ϕ_{lens}	-0.92 (-52°)	Converted from (e_1, e_2)	(c)

Table 6. The true values and priors of the source and lens light parameters used for simulation and fitting of the lensed quasar system.

(a) The lensed positions are sampled first and their unlensed position was evaluated. (b) $q = \frac{1-c}{1+c}$, $c = \sqrt{e_1^2 + e_2^2}$. (c) $\phi = \frac{1}{2} \arctan 2(e_2, e_1)$

Definitions. The axis ratio $q = b/a$ ($0 < q \leq 1$) and major-axis position angle $\phi \in [0, \pi]$ are mapped to the Cartesian ellipticity components¹¹

$$c = \frac{1-q}{1+q}, \quad e_1 = c \cos 2\phi, \quad e_2 = c \sin 2\phi. \quad (\text{E.1})$$

Sampling in the (e_1, e_2) avoids the discontinuity at $\phi=0=\pi$ and helps a continuous angular sampling, but introduces a *non-uniform* Jacobian with respect to (q, ϕ) :

$$\det \left(\frac{\partial(e_1, e_2)}{\partial(q, \phi)} \right) = \begin{vmatrix} \frac{\partial e_1}{\partial q} & \frac{\partial e_1}{\partial \phi} \\ \frac{\partial e_2}{\partial q} & \frac{\partial e_2}{\partial \phi} \end{vmatrix} = -\frac{4(1-q)}{(1+q)^3} = -c(1+c)^2. \quad (\text{E.2})$$

A flat prior in (e_1, e_2) therefore induces a strong preference for smaller ellipticities ($q < 0.5$) and cannot reach $q < 3 - 2\sqrt{2} \approx 0.17$ with any angle and $3 - 2\sqrt{2} \leq q < 1/3$ for some angles, if the usual box $|e_1|, |e_2| \leq 0.5$ is enforced. Figure 9 (a, c, e) illustrates the bias.

Desired prior. For our lens population we wish to keep $P(q) \propto \text{const}$ and $P(\phi) \propto \text{const}$ before giving the multipole-related prior. Combining Eq. (E.2) with $P(q, \phi) = P(e_1, e_2) |\det J|$

¹¹This is the LENSTRONOMY convention.

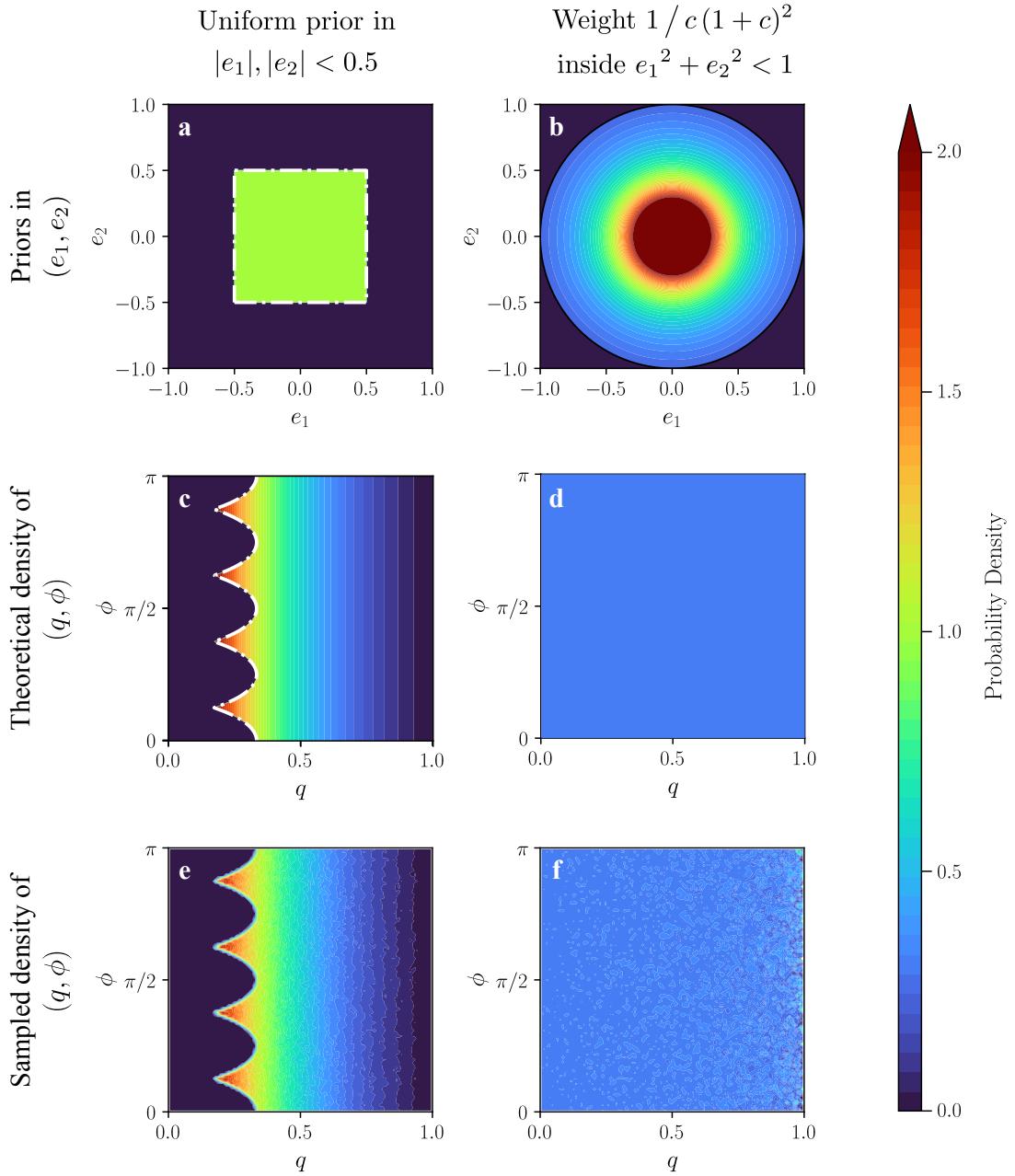


Figure 9. **Top:** Priors in the (e_1, e_2) space. **(a)** Uniform in the square $|e_1|, |e_2| \leq 0.5$. **(b)** Weighted in the unit disk ($c < 1$) by $1/c(1+c)^2$.

Middle: Analytic mapping to (q, ϕ) . **(c)** Flat square in (e_1, e_2) maps to a biased (q, ϕ) prior, increasing sharply as $q \rightarrow 0$, with a hard cutoff at $q > 3 - 2\sqrt{2}$. The white dash-dot line traces the image of the square boundary from panel (a). **(d)** Applying the weight $\frac{1}{c(1+c)^2}$ flattens the prior over (q, ϕ) .

Bottom: Monte Carlo validation. **(e)** Sampling from the square reproduces (c). **(f)** Disk sampling with weight recovers uniformity as in (d). The granular structure at high q results partly from limited number of sampling ($N = 5 \times 10^6$). Additionally, near $q \sim 1$ (i.e., $e_1, e_2 \sim 0$), numerical precision limits in (e_1, e_2) introduce artifacts in the transformed (q, ϕ) space.

yields

$$P(e_1, e_2) \propto \frac{(1+q)^3}{4(1-q)} = \frac{1}{c(1+c)^2} \quad (\text{E.3})$$

The prior applies the entire unit disk $e_1^2 + e_2^2 < 1$, allowing axis ratios down to $q = 0$. We implement this as an additive custom log-prior in the LENSTRONOMY likelihood:

$$\ln \mathcal{L}_e = -\ln[c(1+c)^2] \quad \text{if } e_1^2 + e_2^2 < 1, \quad (\text{E.4})$$

$$\ln \mathcal{L}_e = -\infty \quad \text{otherwise.} \quad (\text{E.5})$$

Figure 9 (b, d, f) illustrates the sampling result with the desired prior.

References

- [1] P.A. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela et al., *Planck 2013 results. xvi. cosmological parameters, Astronomy & Astrophysics* **571** (2014) A16.
- [2] S. Vegetti, S. Birrer, G. Despali, C. Fassnacht, D. Gilman, Y. Hezaveh et al., *Strong gravitational lensing as a probe of dark matter, arXiv preprint arXiv:2306.11781* (2023) .
- [3] A. Bonaca, D.W. Hogg, A.M. Price-Whelan and C. Conroy, *The spur and the gap in gd-1: Dynamical evidence for a dark substructure in the milky way halo, The Astrophysical Journal* **880** (2019) 38.
- [4] J.D. Simon, *The faintest dwarf galaxies, Annual Review of Astronomy and Astrophysics* **57** (2019) 375.
- [5] S. Mao and P. Schneider, *Evidence for substructure in lens galaxies?, Monthly Notices of the Royal Astronomical Society* **295** (1998) 587.
- [6] N. Dalal and C.S. Kochanek, *Direct detection of cold dark matter substructure, The Astrophysical Journal* **572** (2002) 25.
- [7] C.S. Kochanek and N. Dalal, *Tests for substructure in gravitational lenses, The Astrophysical Journal* **610** (2004) 69.
- [8] P.L. Schechter and J. Wambsganss, *Quasar microlensing at high magnification and the role of dark matter: Enhanced fluctuations and suppressed saddle points, The Astrophysical Journal* **580** (2002) 685–695.
- [9] D. Xu, D. Sluse, L. Gao, J. Wang, C. Frenk, S. Mao et al., *How well can cold dark matter substructures account for the observed radio flux-ratio anomalies, Monthly Notices of the Royal Astronomical Society* **447** (2015) 3189.
- [10] A.M. Nierenberg, T. Treu, S.A. Wright, C.D. Fassnacht and M.W. Auger, *Detection of substructure with adaptive optics integral field spectroscopy of the gravitational lens B1422+231, Monthly Notices of the Royal Astronomical Society* **442** (2014) 2434.
- [11] A.B. Congdon and C.R. Keeton, *Multipole models of four-image gravitational lenses with anomalous flux ratios, Monthly Notices of the Royal Astronomical Society* **364** (2005) 1459.
- [12] J.W. Hsueh, W. Enzi, S. Vegetti, M.W. Auger, C.D. Fassnacht, G. Despali et al., *SHARP – VII. New constraints on the dark matter free-streaming properties and substructure abundance from gravitationally lensed quasars, Monthly Notices of the Royal Astronomical Society* **492** (2019) 3047.

- [13] D. Gilman, S. Birrer, A. Nierenberg, T. Treu, X. Du and A. Benson, *Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with eight quadruple-image strong gravitational lenses*, *MNRAS* **491** (2020) 6077.
- [14] D. Gilman, S. Birrer, T. Treu, A. Nierenberg and A. Benson, *Probing dark matter structure down to 107 solar masses: flux ratio statistics in gravitational lenses with line-of-sight haloes*, *Monthly Notices of the Royal Astronomical Society* **487** (2019) 5721.
- [15] A.M. Nierenberg, T. Treu, G. Brammer, A.H.G. Peter, C.D. Fassnacht, C.R. Keeton et al., *Probing dark matter substructure in the gravitational lens HE 0435-1223 with the WFC3 grism*, *Monthly Notices of the Royal Astronomical Society* **471** (2017) 2224.
- [16] A.J. Shajib, S. Birrer, T. Treu, M. Auger, A. Agnello, T. Anguita et al., *Is every strong lens model unhappy in its own way? uniform modelling of a sample of 13 quadruply+ imaged quasars*, *Monthly Notices of the Royal Astronomical Society* **483** (2019) 5649.
- [17] C.M. O’Riordan, S.J. Warren and D.J. Mortlock, *Galaxy mass profiles from strong lensing i: the circular power-law model*, *Monthly Notices of the Royal Astronomical Society* **487** (2019) 5143
[<https://academic.oup.com/mnras/article-pdf/487/4/5143/28884393/stz1603.pdf>].
- [18] T. Schmidt, T. Treu, S. Birrer, A.J. Shajib, C. Lemon, M. Millon et al., *Strides: automated uniform models for 30 quadruply imaged quasars*, *Monthly Notices of the Royal Astronomical Society* **518** (2023) 1260.
- [19] J.-W. Hsueh, C.D. Fassnacht, S. Vegetti, J.P. McKean, C. Spingola, M.W. Auger et al., *SHARP – II. Mass structure in strong lenses is not necessarily dark matter substructure: a flux ratio anomaly from an edge-on disc in B1555+375*, *Monthly Notices of the Royal Astronomical Society: Letters* **463** (2016) L51.
- [20] D. Gilman, S. Birrer, A. Nierenberg and M.S.H. Oh, *Turbocharging constraints on dark matter substructure through a synthesis of strong lensing flux ratios and extended lensed arcs*, 2024.
- [21] H.R. Stacey, D.M. Powell, S. Vegetti, J.P. McKean, C.D. Fassnacht, D. Wen et al., *Complex angular structure of three elliptical galaxies from high-resolution alma observations of strong gravitational lenses*, 2024.
- [22] C. Hao, S. Mao, Z. Deng, X. Xia and H. Wu, *Isophotal shapes of elliptical/lenticular galaxies from the Sloan Digital Sky Survey*, *Monthly Notices of the Royal Astronomical Society* **370** (2006) 1339.
- [23] C.S. Kochanek, *The implications of lenses for galaxy structure, Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 373, June 1, 1991, p. 354-368*. **373** (1991) 354.
- [24] C.S. Trotter, J.N. Winn and J.N. Hewitt, *A multipole-taylor expansion for the potential of the gravitational lens mg j0414+ 0534*, *The Astrophysical Journal* **535** (2000) 671.
- [25] O. Möller, P. Hewett and A. Blain, *Discs in early-type lensing galaxies: effects on magnification ratios and measurements of h 0*, *Monthly Notices of the Royal Astronomical Society* **345** (2003) 1.
- [26] N.W. Evans and H.J. Witt, *Fitting gravitational lenses: truth or delusion*, *Monthly Notices of the Royal Astronomical Society* **345** (2003) 1351.
- [27] D. Gilman, A. Agnello, T. Treu, C.R. Keeton and A.M. Nierenberg, *Strong lensing signatures of luminous structure and substructure in early-type galaxies*, *Monthly Notices of the Royal Astronomical Society* **467** (2017) 3970.
- [28] L. Van de Vyvere, Gomer, Matthew R., Sluse, Dominique, Xu, Dandan, Birrer, Simon, Galan, Aymeric et al., *Tdcosmo - vii. boxyness/discyness in lensing galaxies: Detectability and impact on h0*, *A&A* **659** (2022) A127.

- [29] A. Etherington, J.W. Nightingale, R. Massey, S.-I. Tam, X. Cao, A. Niemiec et al., *Strong gravitational lensing's 'external shear' is not shear*, 2023.
- [30] D. Gilman, J. Bovy, T. Treu, A. Nierenberg, S. Birrer, A. Benson et al., *Strong lensing signatures of self-interacting dark matter in low-mass haloes*, *MNRAS* **507** (2021) 2432.
- [31] A. Laroche, D. Gilman, X. Li, J. Bovy and X. Du, *Quantum fluctuations masquerade as haloes: bounds on ultra-light dark matter from quadruply imaged quasars*, *MNRAS* **517** (2022) 1867.
- [32] D. Gilman, A. Benson, J. Bovy, S. Birrer, T. Treu and A. Nierenberg, *The primordial matter power spectrum on sub-galactic scales*, *MNRAS* **512** (2022) 3163.
- [33] D. Gilman, Y.-M. Zhong and J. Bovy, *Constraining resonant dark matter self-interactions with strong gravitational lenses*, *Phys. Rev. D.* **107** (2023) 103008.
- [34] R.E. Keeley, A.M. Nierenberg, D. Gilman, C. Gannon, S. Birrer, T. Treu et al., *Jwst lensed quasar dark matter survey ii: Strongest gravitational lensing limit on the dark matter free streaming length to date*, 2024.
- [35] A.J. Shajib, S. Birrer, T. Treu, A. Agnello, E. Buckley-Geer, J. Chan et al., *Strides: a 3.9 per cent measurement of the hubble constant from the strong lens system des j0408- 5354*, *Monthly Notices of the Royal Astronomical Society* **494** (2020) 6072.
- [36] A.M. Nierenberg, D. Gilman, T. Treu, G. Brammer, S. Birrer, L. Moustakas et al., *Double dark matter vision: twice the number of compact-source lenses with narrow-line lensing and the WFC3 grism*, *Monthly Notices of the Royal Astronomical Society* **492** (2019) 5314.
- [37] A. Nierenberg, R. Keeley, D. Sluse, D. Gilman, S. Birrer, T. Treu et al., *Jwst lensed quasar dark matter survey i: Description and first results*, *arXiv preprint arXiv:2309.10101* (2023) .
- [38] S. Birrer and A. Amara, *lenstronomy: Multi-purpose gravitational lens modelling software package*, *Physics of the Dark Universe* **22** (2018) 189.
- [39] S. Birrer, A.J. Shajib, D. Gilman, A. Galan, J. Aalbers, M. Millon et al., *lenstronomy ii: A gravitational lensing software ecosystem*, *arXiv preprint arXiv:2106.05976* (2021) .
- [40] N. Tessore and R.B. Metcalf, *The elliptical power law profile lens*, *Astronomy & Astrophysics* **580** (2015) A79.
- [41] B.C. Ciambur, *Beyond ellipse(s): Accurately modeling the isophotal structure of galaxies with isofit and cmodel*, *The Astrophysical Journal* **810** (2015) 120.
- [42] H. Paugnat and D. Gilman, *Elliptical multipoles for gravitational lenses*, 2025.
- [43] C.R. Keeton, B.S. Gaudi and A.O. Petters, *Identifying lenses with small-scale structure. i. cusp lenses*, *The Astrophysical Journal* **598** (2003) 138.
- [44] P. Schneider and C. Seitz, *Steps towards nonlinear cluster inversion through gravitational distortions. i. basic considerations and circular clusters*, *arXiv preprint astro-ph/9407032* (1994) .
- [45] P. Schneider, C. Kochanek and J. Wambsganss, *Gravitational lensing: strong, weak and micro: Saas-Fee advanced course 33*, vol. 33, Springer Science & Business Media (2006).
- [46] M. Meneghetti, *Introduction to gravitational lensing: with Python examples*, vol. 956, Springer Nature (2021).
- [47] J. Sérsic, *Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy*, *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina* **6** (1963) 41.
- [48] B. Allgood, R.A. Flores, J.R. Primack, A.V. Kravtsov, R.H. Wechsler, A. Faltenbacher et al., *The shape of dark matter haloes: dependence on mass, redshift, radius and formation*, *Monthly Notices of the Royal Astronomical Society* **367** (2006) 1781
[<https://academic.oup.com/mnras/article-pdf/367/4/1781/4902633/mnras0367-1781.pdf>].

- [49] G. Despali, C. Giocoli and G. Tormen, *Some like it triaxial: the universality of dark matter halo shapes and their evolution along the cosmic time*, *Monthly Notices of the Royal Astronomical Society* **443** (2014) 3208 [<https://academic.oup.com/mnras/article-pdf/443/4/3208/6292768/stu1393.pdf>].
- [50] M.W. Auger, T. Treu, A.S. Bolton, R. Gavazzi, L.V.E. Koopmans, P.J. Marshall et al., *The sloan lens acs survey. ix. colors, lensing, and stellar masses of early-type galaxies*, *The Astrophysical Journal* **705** (2009) 1099.
- [51] M. Barnabè, O. Czoske, L.V.E. Koopmans, T. Treu and A.S. Bolton, *Two-dimensional kinematics of slacs lenses – iii. mass structure and dynamics of early-type lens galaxies beyond $z \simeq 0.1$* , *Monthly Notices of the Royal Astronomical Society* **415** (2011) 2215 [<https://academic.oup.com/mnras/article-pdf/415/3/2215/17328519/mnras0415-2215.pdf>].
- [52] A. Rest, F.C. van den Bosch, W. Jaffe, H. Tran, Z. Tsvetanov, H.C. Ford et al., *Wfpc2 images of the central regions of early-type galaxies. i. the data*, *The Astronomical Journal* **121** (2001) 2431.
- [53] A. Pasquali, I. Ferreras, N. Panagia, E. Daddi, S. Malhotra, J.E. Rhoads et al., *The structure and star formation history of early-type galaxies in the ultra deep field/grapes survey*, *The Astrophysical Journal* **636** (2006) 115.
- [54] K. Mitsuda, M. Doi, T. Morokuma, N. Suzuki, N. Yasuda, S. Perlmutter et al., *Isophote shapes of early-type galaxies in massive clusters at $z = 1$ and 0*, *The Astrophysical Journal* **834** (2017) 109.
- [55] Q. He, J.W. Nightingale, A. Amyrosiadis, A. Robertson, S. Cole, C.S. Frenk et al., *Unveiling lens light complexity with a novel multi-gaussian expansion approach for strong gravitational lensing*, 2024.
- [56] R. Narayan and M. Bartelmann, *Lectures on gravitational lensing*, arXiv preprint [astro-ph/9606001](https://arxiv.org/abs/astro-ph/9606001) (1996) .
- [57] A. Agnello, H. Lin, N. Kuropatkin, E. Buckley-Geer, T. Anguita, P.L. Schechter et al., *DES meets Gaia: discovery of strongly lensed quasars from a multiplet search*, *Monthly Notices of the Royal Astronomical Society* **479** (2018) 4345.
- [58] A. Krone-Martins, L. Delchambre, O. Wertz, C. Ducourant, F. Mignard, R.T.J. Klüter et al., *Gaia gral: Gaia dr2 gravitational lens systems. i. new quadruply imaged quasar candidates around known quasars*, *Astronomy and Astrophysics-A&A* **616** (2018) id.
- [59] D. Gilman, S. Birrer, T. Treu, C.R. Keeton and A. Nierenberg, *Probing the nature of dark matter by forward modelling flux ratios in strong gravitational lenses*, *Monthly Notices of the Royal Astronomical Society* **481** (2018) 819.
- [60] L. Van de Vyvere, D. Sluse, S. Mukherjee, D. Xu and S. Birrer, *The impact of mass map truncation on strong lensing simulations*, *Astronomy & Astrophysics* **644** (2020) A108.
- [61] R. Bender, S. Doebeleiner and C. Moellenhoff, *Isophote shapes of elliptical galaxies. i-the data*, *Astronomy and Astrophysics Supplement Series (ISSN 0365-0138)*, vol. 74, no. 3, Sept. 1988, p. 385-426. **74** (1988) 385.