SHARP - III. First use of adaptive-optics imaging to constrain cosmology with gravitational lens time delays

Geoff C. -F. Chen, Sherry H. Suyu, Kenneth C. Wong, Christopher D. Fassnacht, Tzihong Chiueh, I. Shing Hu, Matthew W. Auger, Léon V. E. Koopmans, David J. Lagattuta, John P. Mckean, et al.

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **462,** 3457–3475 (2016)
Advance Access publication 2016 August 15

doi:10.1093/mnras/stw991

SHARP – III. First use of adaptive-optics imaging to constrain cosmology with gravitational lens time delays

Geoff C.-F. Chen,[1,2,3★] Sherry H. Suyu,[1,4★] Kenneth C. Wong,[1,5★] Christopher D. Fassnacht,[3] Tzihong Chiueh,[2,6,7] Aleksi Halkola, I Shing Hu,[8] Matthew W. Auger,[9] Léon V. E. Koopmans,[10] David J. Lagattuta,[11] John P. McKean[10,12] and Simona Vegetti[4]

[1]*Institute of Astronomy and Astrophysics, Academia Sinica, PO Box 23-141, Taipei 10617, Taiwan*
[2]*Department of Physics, National Taiwan University, Taipei 10617, Taiwan*
[3]*Department of Physics, University of California, Davis, CA 95616, USA*
[4]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, D-85740 Garching, Germany*
[5]*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*
[6]*Institute of Astrophysics, National Taiwan University, Taipei 10617, Taiwan*
[7]*Center for Theoretical Sciences, National Taiwan University, Taipei 10617, Taiwan*
[8]*Department of Mathematics, National Taiwan University, Taipei 10617, Taiwan*
[9]*Institute of Astronomy, University of Cambridge, Madingley Rd, Cambridge CB3 0HA, UK*
[10]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[11]*CRAL, Observatoire de Lyon, Universit Lyon 1, 9 Avenue Ch. Andr, F-69561 Saint Genis Laval Cedex, France*
[12]*Netherlands Institute for Radio Astronomy (ASTRON), PO Box 2, NL-7990 AA Dwingeloo, the Netherlands*

Accepted 2016 April 25. Received 2016 April 20; in original form 2015 December 14

ABSTRACT

Accurate and precise measurements of the Hubble constant are critical for testing our current standard cosmological model and revealing possibly new physics. With *Hubble Space Telescope* (*HST*) imaging, each strong gravitational lens system with measured time delays can allow one to determine the Hubble constant with an uncertainty of ∼7 per cent. Since *HST* will not last forever, we explore adaptive-optics (AO) imaging as an alternative that can provide higher angular resolution than *HST* imaging but has a less stable point spread function (PSF) due to atmospheric distortion. To make AO imaging useful for time-delay-lens cosmography, we develop a method to extract the unknown PSF directly from the imaging of strongly lensed quasars. In a blind test with two mock data sets created with different PSFs, we are able to recover the important cosmological parameters (time-delay distance, external shear, lens-mass profile slope, and total *Einstein* radius). Our analysis of the Keck AO image of the strong lens system RXJ 1131−1231 shows that the important parameters for cosmography agree with those based on *HST* imaging and modelling within 1σ uncertainties. Most importantly, the constraint on the model time-delay distance by using AO imaging with 0.09 arcsec resolution is tighter by ∼50 per cent than the constraint of time-delay distance by using *HST* imaging with 0.09 arcsec when a power-law mass distribution for the lens system is adopted. Our PSF reconstruction technique is generic and applicable to data sets that have multiple nearby point sources, enabling scientific studies that require high-precision models of the PSF.

Key words: gravitational lensing: strong – instrumentation: adaptive optics – methods: data analysis – distance scale.

1 INTRODUCTION

The discovery of the accelerated expansion of the Universe (Riess et al. 1998; Perlmutter et al. 1999) and observations of the cosmic microwave background (CMB; e.g. Hinshaw et al. 2013; Planck Collaboration XIII 2015) have established a standard cosmological

★ E-mail: chfchen@ucdavis.edu (GC-FC); suyu@mpa-garching.mpg.de (SHS); kcwong@asiaa.sinica.edu.tw (KCW)

paradigm where our Universe is spatially flats and is dominated by cold dark matter (CDM) and dark energy: the so-called flat ΛCDM model, where Λ represents a constant dark energy density. While the CMB provides strong constraints on the parameters of this model, a relaxation of the assumptions in this model, such as spatial flatness or constant dark energy density, leads to a strong degeneracy between the cosmological parameters, particularly those with the Hubble constant H_0. While combining baryon acoustic oscillations/supernovae/weak lensing with the CMB partially removes the need to make those assumptions (e.g. Linder 2004; Coe & Moustakas 2009; Suyu et al. 2013; Planck Collaboration XIII 2015), Jee et al. (2016) show that the combined lensing information significantly helps to constrain cosmological parameters, particularly when curvature is allowed to vary and when the equation of state of dark energy is allowed to be time dependent. Furthermore, when compared to the space-based observational efforts to calibrate Cepheids and Type Ia supernovae, measuring quasar time delays is relatively inexpensive (e.g. Tewes, Courbin & Meylan 2013a). Therefore, independent and accurate measurements of H_0 provide useful complements to the observations of the CMB in constraining the spatial curvature of the Universe, dark energy equation of state, and the number of neutrino species (e.g. Hu 2005; Riess et al. 2009, 2011; Freedman et al. 2012; Suyu et al. 2012a). The recent inferred value of Hubble constant $H_0 = 67.8 \pm 0.9\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, based on the *Planck* satellite data of the CMB and the assumption of the flat ΛCDM model, is low in comparison to several direct measurements including those from the Cepheids distance ladder with $H_0 = 74.3 \pm 1.5(\mathrm{stat.}) \pm 2.1(\mathrm{sys.})\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Freedman et al. 2012) and $H_0 = 73.8 \pm 2.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Riess et al. 2011). If this indication of tension is not ruled out by systematic effects, then this could indicate new physics beyond the standard flat ΛCDM model. Therefore, 'pinning down' Hubble constant to 1 per cent precision with independent techniques is critical for better understanding our Universe (Suyu et al. 2012a).

Strong gravitational lensing with time delays provides a one-step measurement of a cosmological distance in the Universe. When compared to the space-based observational efforts to calibrate Cepheids and Type Ia supernovae, measuring time delays and obtaining follow-up imaging/spectroscopy of strong lens systems are relatively inexpensive (Suyu et al. 2013; Tewes et al. 2013a, e.g.). In a time-delay lens, the background source is composed of a centrally varying source, such as an active galactic nucleus (AGN), and its host galaxy. The time delays between the multiple images of the source, induced by the foreground lens, are given by $\Delta t = \frac{1}{c} D_{\Delta t} \Delta\tau$. Here, $D_{\Delta t}$ is the time-delay distance that encompasses cosmological dependences and is particularly sensitive to the Hubble constant (e.g. Suyu et al. 2010) and $\Delta\tau$ is dependent on the geometry and the gravitational potential of the lens system; $\Delta\tau$ can be tightly constrained by the spatially extended images (usually known as 'arcs') of the lensed background galaxy (e.g. Kochanek, Keeton & McLeod 2001; Suyu et al. 2009), together with stellar kinematics of the foreground lens galaxy (e.g. Treu & Koopmans 2002; Koopmans et al. 2003; Suyu et al. 2010, 2014) and studies of the lens environment combined with ray-tracing through numerical simulations (e.g. Hilbert et al. 2007, 2009; Suyu et al. 2010; Fassnacht, Koopmans & Wong 2011; Collett et al. 2013; Greene et al. 2013). The stellar kinematics and lens environment studies are important for overcoming the mass-sheet degeneracy and source-position transformations in lensing (Falco, Gorenstein & Shapiro 1985; Schneider & Sluse 2013, 2014; Xu et al. 2016). Therefore, by measuring the time delays between the multiple images and modelling the lens and line-of-sight mass distributions, we can constrain $D_{\Delta t}$. The time delays in combination with the stellar velocity dispersion measurements of the lens galaxy further allow us to infer the angular diameter distance to the lens galaxy (Paraficz & Hjorth 2009; Jee, Komatsu & Suyu 2015).

Suyu et al. (2013) have shown that for each lens system we can measure H_0 to ∼7 per cent precision. *Hubble Space Telescope* (*HST*) imaging is imperative for this analysis because it not only provides high angular resolution but also a stable point spread function (PSF) for the lens-mass modelling. However, *HST*'s lifetime is finite,[1] and the angular resolution is also limited by its aperture size. Given the dozens of time-delay lenses from COSMOGRAIL[2] (e.g. Vuissoz et al. 2007, 2008; Courbin et al. 2011; Eulaers et al. 2013; Rathna Kumar et al. 2013; Tewes et al. 2013a,b), and hundreds of new lenses to be discovered in the near future (e.g. Oguri & Marshall 2010; Agnello et al. 2015; Chan et al. 2015; Marshall et al. 2016; More et al. 2016), finding an alternative long-term solution for this promising method is timely.

One alternative approach is imaging from the ground via adaptive optics (AO), which is a technology used to improve the performance of optical systems by reducing the effect of wavefront distortions (e.g. Rousset et al. 1990; Beckers 1993; Watson 1997; Brase 1998). In other words, it aims at correcting the deformations of an incoming wavefront by deforming a mirror and thus compensating for the distortion. The advantages of using AO imaging are (1) the angular resolution obtained with telescopes that are larger than *HST* can be higher than that of *HST* since a perfect AO system would lead to a diffraction-limited PSF, (2) ground-based telescopes are more accessible. The disadvantage is that we do not have a stable PSF model a priori, since the atmospheric distortion varies both temporally and spatially across the image. Lens targets typically do not have a nearby bright star within ∼10 arcsec, and stars at further angular distance from the target may be insufficient in providing an accurate PSF model given the spatial variation of the PSF across the field.

In *HST* imaging, we can use the lensing arcs to constrain the lens-mass model by using the stable PSF of *HST* to separate the arc from the bright AGN. The contamination of the AGN light on the lensing arcs in AO imaging makes it difficult to constrain the lens model, and consequently H_0. One therefore needs to obtain a good PSF model for the AO data, and there are recent studies that aim to do so directly from the AO imaging. Lagattuta, Auger & Fassnacht (2010) use three Gaussian components as the PSF model to subtract the AGN light which is sufficient to study the lensing galaxy and its substructures. However, the analytical model is not sufficient to describe the complexity of the PSF (see fig. 1 of Lagattuta et al. 2010) which could potentially impact the cosmographic measurements. Rusu et al. (2016) use either an analytic or a hybrid PSF to study the host galaxies of the lensed AGNs (see also Rusu et al. 2014). The hybrid PSF is built from elliptical Moffat profiles (Moffat 1969) with central parts iteratively tuned to match a single AGN image. While this hybrid PSF is useful for extracting properties of the AGN host galaxy, the central parts of the PSF model could manifest the noise pattern in the image (see fig. B.7 of Rusu et al. 2016) which also could potentially impact cosmographic measurements. Agnello et al. (2015) use an iterative method to reconstruct the PSF directly from lens imaging by averaging the doubly lensed AGN. This method is valid only when the lensed AGN are separated far

[1] And no equivalent optical space-based telescope might be forthcoming soon.

[2] COSmological MOnitoring of GRAvItational Lenses.

enough from each other. Chantry & Magain (2007) use the MCS deconvolution algorithm (Magain, Courbin & Sohy 1998) to reconstruct the PSF and the lensing arcs, but did not incorporate/quantify a lens mass model in their approach. For typical quad (four-image) lens systems, the lensed AGNs are often close in separation (within 2 arcsec), leading to overlaps in the wings of the AGN images that are smeared by the PSF.

Our goal is to provide a general method to overcome the unknown PSF model problem by extracting the PSF directly from strong-lensing imaging and simultaneously modelling the lens-mass distribution. We develop a method that accounts for such overlaps and reconstruct simultaneously the PSF, the lensing arcs of the AGN host galaxy and the lens-mass distribution. We test our method on simulated AO images, and apply the method to the known gravitational lens RXJ 1131−1231 with Keck AO imaging, a part of data from SHARP,[3] which is a project that focuses on studying known quadruple-image and *Einstein* ring lenses using high-resolution AO imaging, in order to probe their mass distributions in unprecedented detail (e.g. Lagattuta et al. 2010, 2012; Vegetti et al. 2012; Hsueh et al. 2016). The gravitational lens system RXJ 1131−1231 was discovered by Sluse et al. (2003) who also measured the lens and source redshifts to be 0.295 and 0.658, respectively. The *HST* observations of the system RXJ 1131−1231 have been modelled by Suyu et al. (2013, 2014) for cosmography and more recently by Birrer, Amara & Refregier (2016).

The outline of the paper is as follows. In Section 2, we describe the observation of RXJ 1131−1231 with the AO imaging system at the Keck Observatory. We briefly recap in Section 3 the basics of cosmography with time-delay lenses. In Section 4, we describe our new procedure to analyse AO images without information on the PSF in advance. In Section 5, we use simulated data to test and verify the method. In Section 6, we demonstrate the results from real data and provide a comparison between the results from *HST* imaging and AO imaging. Finally, we summarize in Section 7.

2 OBSERVATION

The RXJ 1131−1231 system was observed on the nights of UT 2012 May 16 and May 18 with the Near-Infrared Camera 2 on the Keck-2 Telescope (e.g. Wizinowich et al. 2003). This image was a part of SHARP data. The AO corrections were achieved through the use of an $R = 15.8$ tip-tilt star located 54.5 arcsec from the lens system and a laser guide star. The system was observed in the 'Wide Camera' mode, which provides a roughly 40arcsec × 40arcsec field of view and a pixel scale of 0.0397 arcsec. This pixel scale slightly undersamples the PSF, but the angular extent of the lens system and the distance from the tip-tilt star made the use of the Wide Camera the preferable approach.

The observations consisted of 61 exposures, each consisting of six co-added 10 s exposures, for a total on-source integration time of 3660 s. The data were reduced by a PYTHON-based pipeline that has steps that do the flat-field correction, subtract the sky, correct for the optical distortions in the raw images, and combine the calibrated data frames (for details, see Auger et al. 2011). The final image has a pixel-scale of 0.04 arcsec and is shown in Fig. 1.

[3] Strong-lensing High Angular Resolution Programme (Fassnacht et al. in preparation)

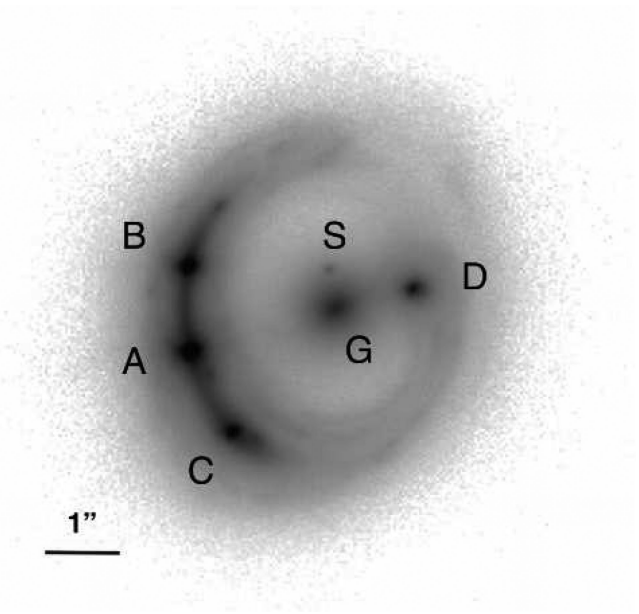

Figure 1. Keck AO image (K′ band) of the gravitational lens RXJ1131−1231. The lensed AGN image of the spiral source galaxy is marked by A, B, C, and D, and the star-forming regions in the background spiral galaxy form plentiful lensed features. The foreground main lens and the satellite are indicated by G and S, respectively.

3 BASIC THEORY

3.1 The theory of gravitational lensing with time delay

In this section, we briefly explain the relation between gravitational time delays and cosmology. When a light ray passes near a massive object, it experiences a deflection in its trajectory and acquires a time delay by the gravitational field with respect to the travel time without the massive object. Therefore, the time delay has two contributions: (1) the geometric delay, $\Delta t_{\rm geom}$, which is caused by the bent trajectory being longer than the unbent one, and (2) the gravitational delay, $\Delta t_{\rm grav}$, which is due to the fact that the space and time are affected around the gravitational field, so after integrating the gravitational potential along the path, a faraway observer receives the light later by a Shapiro delay (Refsdal 1964; Shapiro 1964).

The combination of the two delays is

$$\Delta t = \frac{D_{\Delta t}}{c}\left[\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta})\right], \quad (1)$$

where $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, and $\psi(\boldsymbol{\theta})$ are the image coordinates, the source coordinates, and the lens potential, respectively. The time-delay distance is defined as

$$D_{\Delta t} \equiv (1+z_{\rm d})\frac{D_{\rm d}D_{\rm s}}{D_{\rm ds}} \propto H_0^{-1}, \quad (2)$$

where $D_{\rm d}$, $D_{\rm s}$, and $D_{\rm ds}$ are the angular diameter distances to the lens, to the source, and between the lens and the source, respectively. Thus, we can measure $D_{\Delta t}$ via gravitational lensing with time delays. Notice that the gradient of the term in the square brackets in equation (1) vanishes at the positions of the lensed images and yields the lens equation

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \nabla\psi(\boldsymbol{\theta}), \quad (3)$$

which governs the deflection of light rays.

We refer the reader to, e.g. Schneider, Kochanek & Wambsganss (2006), Bartelmann (2010), Treu (2010), Suyu et al. (2010), and Treu & Ellis (2015) for more details.

3.2 Probability theory

A meaningful measurement should have an uncertainty as a reference and it is also the key to confirm or rule out possible models. Thus, we need to analyse our data based on a probability theory that can present this idea. Bayes' theorem provides the conditional probability distribution, so we can obtain the posterior probability distribution of the model parameters given the data from Bayes' rule. For example, if we are interested in the posterior of the parameters π of the hypothesis model H given the data $\boldsymbol{d}$, it can be expressed as

$$\overbrace{P(\pi|\boldsymbol{d},H)}^{\text{posterior}} = \frac{\overbrace{P(\boldsymbol{d}|\pi,H)}^{\text{likelihood}}\overbrace{P(\pi|H)}^{\text{prior}}}{\underbrace{P(\boldsymbol{d}|H)}_{\text{evidence(marginalizedlikelihood)}}}, \tag{4}$$

where the Bayesian evidence can be used to rank the model and our prior based on the data (e.g. MacKay 1992; Hobson, Bridle & Lahav 2002; Marshall et al. 2002)

In addition, if we are interested in the posterior of a specific parameter, π_N, the posterior distribution can be obtained by marginalizing over other parameters

$$P(\pi_N|\boldsymbol{d},H) = \int P(\pi|\boldsymbol{d},H)\prod_{i=1}^{N-1} \mathrm{d}\pi_i. \tag{5}$$

3.3 Markov Chain Monte Carlo

Obtaining the probability distribution function of the parameters in a model can be non-trivial, especially when the number of parameters is high. It is computationally unfeasible to explore a high-dimensional parameter space on a regular grid since the number of the grid points for the task exponentially increases with the number of dimensions. Due to the fact that the parameter space is typically large in strong-lensing analyses, one can bypass the use of grids by obtaining samples in the multidimensional parameter space that represent the probability distribution (i.e. the number density of the samples is proportional to the probability density). A Markov Chain Monte Carlo (MCMC) provides an efficient way to draw samples from the posterior probability density function (PDF) of the lens parameters, because of the approximately linear relation between the computational time and the dimension of the parameter space.

We use MCMC sampling that is implemented in GLEE, a strong lens modelling software developed by S. H. Suyu and A. Halkola (Suyu & Halkola 2010; Suyu et al. 2012b). It is based on Bayes' theorem and follows Dunkley et al. (2005) to achieve efficient sampling and to test convergence. The pragmatic procedure for convergence is described in Suyu & Halkola (2010). We use Bayesian language in the following sections.

4 METHOD: PSF RECONSTRUCTION AND LENS MODELLING

In this section, we describe a novel procedure to analyse the AO imaging without a PSF model a priori. Readers who are not planning to use this method may wish to proceed directly to Section 5 on the scientific results enabled by the method.

The assumption of this strategy is that the PSF does not change significantly within several arcseconds, which is typically valid in AO imaging (van Dam et al. 2006; Wizinowich et al. 2006). We show an overall flow chart in Fig. 2 to illustrate how to obtain iteratively the PSF, background source intensity, the lens-mass and light model.

In Section 4.1, we decompose the observed light from the lens system into three components (lens galaxy, lensed arcs of the background source galaxy, and the lensed background AGN) and introduce the notation that we will use in the subsequent discussion. In Section 4.2, we obtain the preliminary global structure of AGN light model, while separating the lens light and arc light. In Section 4.3, we obtain the fine structure of the AGN light and incorporate it into the preliminary AGN light model. This is accomplished by correcting the PSF model. In Section 4.4, we update the PSF and use it to model the lens-mass and source intensity distributions. Since the lens galaxy light is quite smooth and less sensitive to the PSF model, we use the PSF built from the AGN light for the lens galaxy light model. The PSF updating and lens-mass modelling are repeated until the corrections to the PSF become insignificant. (See the criteria in Sections 4.3.3 4.4.3.)

4.1 Light components of the lens system

As shown in Fig. 3, our model for the observed light in the lens system on the image plane has three contributions: the lens galaxy light, the arc light (the lensed background source, i.e. the host galaxy of the AGN), and the light of the multiple AGNs on the image plane. We define

$$\boldsymbol{d} = \boldsymbol{d}^{\mathrm{P}} + \boldsymbol{n}, \tag{6}$$

where $\boldsymbol{d}$ is the vector of observed data (image pixel values),

$$\boldsymbol{d}^{\mathrm{P}} = \overbrace{\mathbf{K}\boldsymbol{g}}^{\text{lenslight}} + \overbrace{\mathbf{K}\mathbf{L}\boldsymbol{s}}^{\text{arclight}} + \overbrace{\mathbf{M}\boldsymbol{w}}^{\text{AGNlight}}, \tag{7}$$

and $\boldsymbol{n}$ is the noise in the data characterized by the covariance matrix $\mathbf{C}_{\mathrm{D}}$ (we use subscript D to indicate 'data'). The blurring matrix $\mathbf{K}$ accounts for the PSF convolution, the vector $\boldsymbol{g}$ is the image pixel values of the lens galaxy light, the matrix $\mathbf{L}$ maps source intensity to the image plane, the vector $\boldsymbol{s}$ describes the source surface brightness on a grid of pixels, the matrix $\mathbf{M}$ is composed using the positions and the intensities of the AGNs, and $\boldsymbol{w}$ is the vector of pixel values of the PSF grid. We refer to Treu & Koopmans (2004) for constructing $\mathbf{K}$ and $\mathbf{L}$, and illustrate the effect of $\mathbf{M}$ in Fig. 3.

At first, since we do not know the AO PSF a priori, $\mathbf{K}$ and $\boldsymbol{w}$ are just the initial blurring matrix and PSF grid values, respectively. As we iteratively model the light components and correct the PSF, we update $\boldsymbol{w}$ (and subsequently $\mathbf{K}$).

4.2 Determining the light components

The goal in this section is to obtain the preliminary model of each of the three light components. In step 1 of Fig. 2, we input the observed image into the lens modelling software GLEE with a nearby star as our initial PSF model. If there is no nearby star, any star in the field can be used as the initial PSF or we can use one of the AGN images. A different initial PSF does not affect the final results, although we note that a good initial PSF would be helpful as they would require fewer iterations of PSF corrections. In step 2, we decompose the

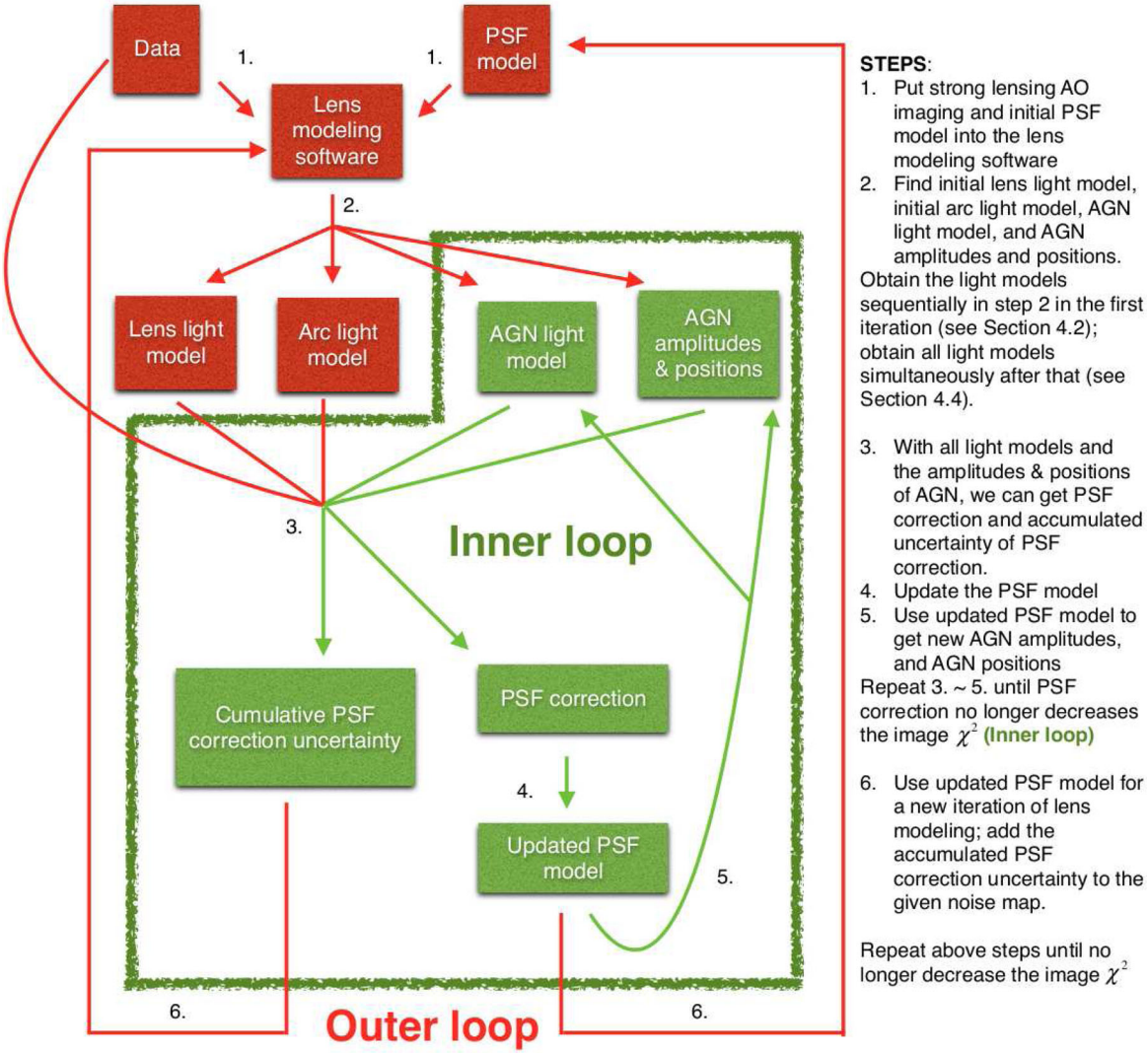

Figure 2. The flow-chart describes the overall procedures in Section 4. We use the procedures to reconstruct the PSF directly from lens image and do the lens modelling. In step 1, we use a nearby star (or one of the lensed AGN itself) as the initial PSF; in step 2, we sequentially obtain the lens light, arc light, AGNs light, and the positions and relative amplitudes of AGNs; steps 3 to 5 form an inner loop to add the correction (fine structures) into the PSF and accumulate the correction uncertainties; in step 6, we enter the outer loop which updates the image covariance matrix, PSF of all light model, and then repeat the full procedure until the image χ^2 no longer decreases.

predicted total light sequentially into lens light, arc light, and AGN light. We detail this process in Section 4.2.1 to Section 4.2.3 below.

4.2.1 Lens light model (step 2)

For modelling the light distribution of the lens galaxy, we use parametrized profiles, such as the elliptical Sérsic profile,

$$I_{\rm S}(\theta_1,\theta_2) = I_{\rm s}\exp\left[-k\left(\left(\frac{\sqrt{\theta_1^2+\theta_2^2/q_{\rm L}^2}}{R_{\rm eff}}\right)^{1/n_{\rm sérsic}}-1\right)\right], \quad (8)$$

where $I_{\rm s}$ is the amplitude, k is a constant such that $R_{\rm eff}$ is the effective radius, $q_{\rm L}$ is the minor-to-major axis ratio, and $n_{\rm sérsic}$ is the Sérsic index (Sérsic 1968).

In order to get a preliminary model of the lens light, we mask out the arc light and AGN light region; that is, we boost the uncertainty of the region where the arc light and the AGN light are apparently dominant. Thus, in the fitting region, equation (7) becomes effectively

$$\boldsymbol{d}^{\rm P} = \mathbf{K}\boldsymbol{g}. \quad (9)$$

By Bayes' rule, we have

$$P(\boldsymbol{\eta}|\boldsymbol{d}) \propto P(\boldsymbol{d}|\boldsymbol{\eta})P(\boldsymbol{\eta}), \quad (10)$$

where $\boldsymbol{\eta}$ represents the parameters of lens light (such as $I_{\rm s}, q_{\rm L}, n_{\rm sérsic}, R_{\rm eff}$). We assume uniform prior on the lens light parameters, so we want to obtain

$$P(\boldsymbol{d}|\boldsymbol{\eta}) = \frac{\exp[-E_{\rm D,mArcAGN}(\boldsymbol{d}|\boldsymbol{\eta})]}{Z_{\rm D,mArcAGN}}, \quad (11)$$

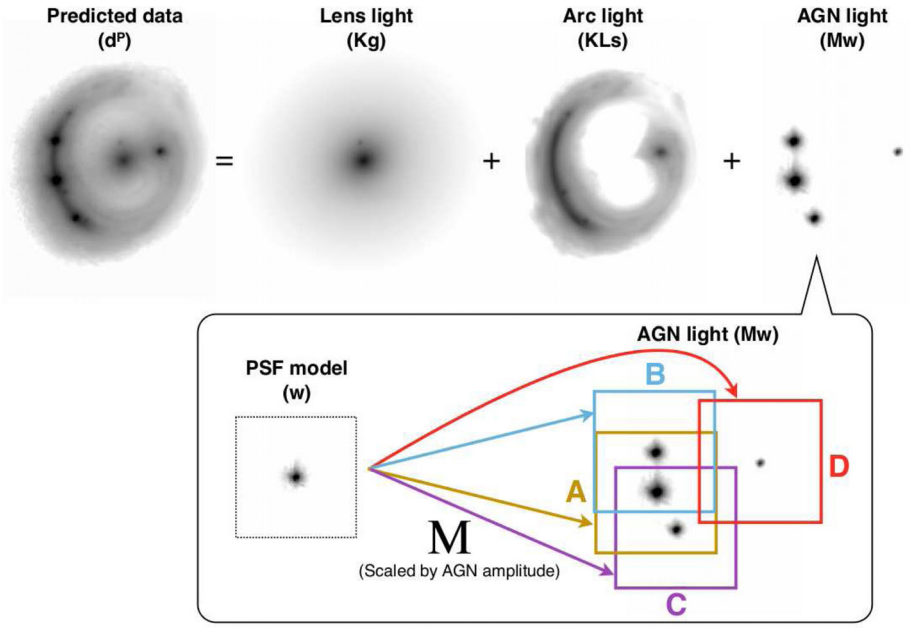

Figure 3. Top panel: we decompose the image into lens light, arc light, and AGN light sequentially. Bottom panel: we model the AGN light by placing the PSF grid (described by vector w) at each of the AGN positions and scaling each PSF by its respective AGN amplitude. This procedure can be characterized by a matrix $\mathbf{M}$, such that the AGN light model on the image plane can be expressed as $\mathbf{M}\boldsymbol{w}$.

where,

$$E_{\mathrm{D,mArcAGN}}(\boldsymbol{d}|\boldsymbol{\eta}) = \frac{1}{2}(\boldsymbol{d}-\mathbf{K}\boldsymbol{g})^{\mathrm{T}}\mathbf{C}^{-1}_{\mathrm{D,mArcAGN}}(\boldsymbol{d}-\mathbf{K}\boldsymbol{g}) = \frac{1}{2}\chi^2_{\mathrm{mArcAGN}}, \quad (12)$$

and

$$Z_{\mathrm{D,mArcAGN}} = (2\pi)^{N_{\mathrm{d}}/2}(\det \mathbf{C}_{\mathrm{D,mArcAGN}})^{1/2} \quad (13)$$

is the normalization for the probability. The covariance matrix, $\mathbf{C}_{\mathrm{D,\,mArcAGN}}$, is the original covariance matrix with entries corresponding to the arc and AGN mask region boosted (see Appendix A), and N_{d} is the number of image pixels. We denote $\hat{\boldsymbol{\eta}}$ as the maximum likelihood parameters [which maximizes equation (11)].

Since the initial PSF is a prototype, usually there are some significant residuals in the lens light centre when maximizing the posterior of lens light parameters. However, this does not affect the subsequent lens light prediction in the arc region, because the residuals are far from the arc regions. To recap, we can obtain $\hat{\boldsymbol{\eta}}$ by masking out the arc light and AGN light regions.

4.2.2 *Arc light model (step 2)*

For modelling the arc light, we describe the source intensity on a grid of pixels on the source plane and map the source intensity values on to the image plane using a lens-mass model [via the operation $\mathbf{KL}\boldsymbol{s}$ in equation (7)]. We use elliptically symmetric power-law distributions to model the dimensionless surface mass density of lens galaxies,

$$\kappa_{\mathrm{pl}}(\theta_1,\theta_2) = \frac{3-\gamma'}{1+q}\left(\frac{\theta_{\mathrm{E}}}{\sqrt{\theta_1^2+\theta_2^2/q^2}}\right)^{\gamma'-1}, \quad (14)$$

where γ' is the radial power-law slope ($\gamma'=2$ corresponding to isothermal), θ_{E} is the *Einstein* radius, and q is the axis ratio of the elliptical isodensity contour. In addition to the lens galaxies, we include a constant external shear with the following lens potential in polar coordinates θ and φ:

$$\psi_{\mathrm{ext}}(\theta,\varphi) = \frac{1}{2}\gamma_{\mathrm{ext}}\theta^2\cos 2(\varphi-\phi_{\mathrm{ext}}), \quad (15)$$

where γ_{ext} is shear strength and ϕ_{ext} is the shear angle. The shear position angle of $\phi_{\mathrm{ext}}=0^\circ$ corresponds to a shearing along θ_1 whereas $\phi_{\mathrm{ext}}=90^\circ$ corresponds to shearing along θ_2.[4]

We model the arc light with the lens light fixed. Since the AGN light dominates near the AGN image positions, we mask out the region where the arc is hard to see; that is, we want to minimize the contribution to the source intensity reconstruction from the AGN light. Since the regions of the AGN are masked out, we temporarily[5] drop the AGN component, $\mathbf{M}\boldsymbol{w}$, in equation (7), which given $\hat{\boldsymbol{\eta}}$ becomes

$$\boldsymbol{d}^{\mathrm{P}} = \mathbf{K}\hat{\boldsymbol{g}} + \mathbf{KL}\boldsymbol{s}, \quad (16)$$

where $\hat{\boldsymbol{g}} = \boldsymbol{g}(\hat{\boldsymbol{\eta}})$. The posterior based on the arc light is

$$P(\boldsymbol{\zeta}|\boldsymbol{d},\Delta\boldsymbol{t},\hat{\boldsymbol{\eta}}) \propto P(\boldsymbol{d},\Delta\boldsymbol{t}|\hat{\boldsymbol{\eta}},\boldsymbol{\zeta})P(\boldsymbol{\zeta}), \quad (17)$$

where $\boldsymbol{\zeta}$ are the parameters of the lens-mass distributions (such as γ', θ_{E}, γ_{ext}). The likelihood of the data can be expressed as

$$P(\boldsymbol{d},\Delta\boldsymbol{t}|\hat{\boldsymbol{\eta}},\boldsymbol{\zeta}) = \int \mathrm{d}\boldsymbol{s}\ \mathrm{P}(\boldsymbol{d},\Delta\boldsymbol{t}|\hat{\boldsymbol{\eta}},\boldsymbol{\zeta},\boldsymbol{s})\mathrm{P}(\boldsymbol{s}), \quad (18)$$

[4] Our (right-handed) coordinate system (θ_1,θ_2) has θ_1 along the east–west direction and θ_2 along the north–south direction.
[5] We will put the AGN component back in next section, 4.2.3.

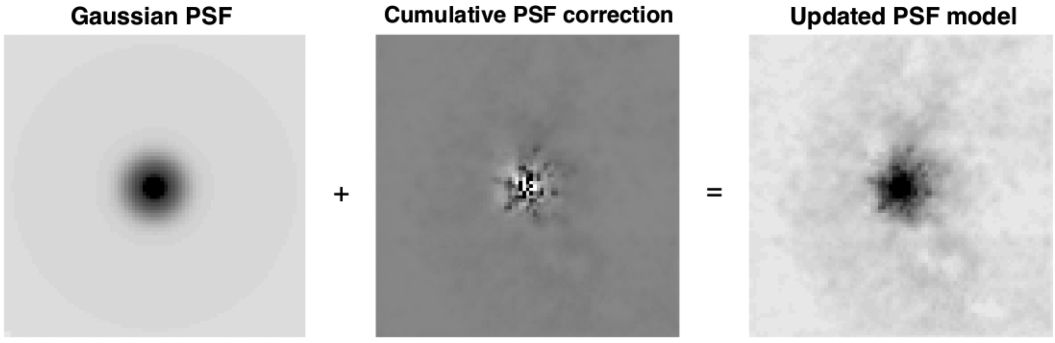

Figure 4. The left-hand panel is the global structure of the PSF. The middle panel is the cumulative fine structure of the PSF. We add the fine structure to global structure to get the PSF model in the right-hand panel.

where

$$P(\boldsymbol{d}, \Delta\boldsymbol{t}|\hat{\boldsymbol{\eta}}, \boldsymbol{\zeta}, \boldsymbol{s}) = \frac{\exp[-E_{\rm D,mAGN}(\boldsymbol{d}|\hat{\boldsymbol{\eta}}, \boldsymbol{\zeta}, \boldsymbol{s})]}{Z_{\rm D,mAGN}} \cdot \prod_{i=1}^{N_{\rm AGN}} \frac{1}{\sqrt{2\pi}\sigma_{{\rm AGN},i}} \exp\left(-\frac{|\boldsymbol{\theta}_{{\rm AGN},i} - \boldsymbol{\theta}^{\rm p}_{{\rm AGN},i}|^2}{2\sigma^2_{{\rm AGN},i}}\right) \cdot \prod_{i=1} \frac{1}{\sqrt{2\pi}\sigma_{\Delta t,i}} \exp\left[-\frac{(\Delta t_i - \Delta t_i^{\rm p})^2}{2\sigma^2_{\Delta t,i}}\right], \quad (19)$$

$$E_{\rm D,mAGN}(\boldsymbol{d}|\hat{\boldsymbol{\eta}}, \boldsymbol{\zeta}, \boldsymbol{s}) = \frac{1}{2}(\boldsymbol{d} - \mathbf{K}\hat{\boldsymbol{g}} - \mathbf{K}\mathbf{L}\boldsymbol{s})^{\rm T}\mathbf{C}^{-1}_{\rm D,mAGN}(\boldsymbol{d} - \mathbf{K}\hat{\boldsymbol{g}} - \mathbf{K}\mathbf{L}\boldsymbol{s}), \quad (20)$$

and

$$Z_{\rm D,mAGN} = (2\pi)^{N_{\rm d}/2}(\det\, \mathbf{C}_{\rm D,mAGN})^{1/2} \quad (21)$$

is the normalization for the probability. We discuss the 'mAGN'regions in Appendix A. In the second term of equation (19), $\boldsymbol{\theta}_{{\rm AGN},i}$ is the measured AGN image position and $\sigma_{{\rm AGN},i}$ is the estimated positional uncertainty of AGN image i; in the third term, Δt_i is the measured time delay with uncertainty $\sigma_{\Delta t,i}$ for image pair $i = AB$, CB, or DB. After we maximize the likelihood of the data, we obtain $\hat{\boldsymbol{\zeta}}$, and also the predicted arc light of the reconstructed background source intensity, $\hat{\boldsymbol{s}}$, of the AGN host galaxy. Note that if there is no time-delay information, one can remove the last term in equation (19).

4.2.3 AGN light model (step 2)

In equation (7), we use $\mathbf{M}\boldsymbol{w}$ to represent the AGN light. In the next section, we further decompose the PSF, $\boldsymbol{w}$, into the global structure and the fine structure that are shown in Fig. 4. In particular, we define

$$\boldsymbol{w} = \boldsymbol{w}_{[0]} + \mathbf{T}_{[0]}\delta\boldsymbol{w}_{[0]}, \quad (22)$$

where $\boldsymbol{w}_{[0]}$ is the vector of global structure, $\delta\boldsymbol{w}_{[0]}$ is the vector of fine structure and the subscript, [0], represents the zeroth iteration. Since, in this section, we focus on the global structure of the PSF, we postpone the discussion of $\mathbf{T}$ to equation (29) and let

$$\boldsymbol{w} = \boldsymbol{w}_{[0]}. \quad (23)$$

By using $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\zeta}}$, and $\hat{\boldsymbol{s}}$ from the previous two sections and keeping them fixed, we model the global structure of the PSF with Gaussian profiles, each of the form

$$I_{\rm G}(\theta_1, \theta_2) = I_{\rm g} \exp\left[-\frac{\theta_1^2 + (\theta_2^2/q_{\rm g}^2)}{2\sigma_{\rm g}^2}\right], \quad (24)$$

where $I_{\rm g}$ is the amplitude, $q_{\rm g}$ is the axis ratio, and $\sigma_{\rm g}$ is the width. We find that a few Gaussians (∼2–4) with a common centroid are sufficient in describing the global structure of the PSF.[6] Substituting equation (23) into equation (7), given $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{s}}$, we obtain

$$\boldsymbol{d}^{\rm P} = \mathbf{K}\hat{\boldsymbol{g}} + \mathbf{K}\hat{\mathbf{L}}\hat{\boldsymbol{s}} + \mathbf{M}\boldsymbol{w}_{[0]}, \quad (25)$$

where $\hat{\mathbf{L}} = \mathbf{L}(\hat{\boldsymbol{\zeta}})$, which is kept fixed at this step. Note that the $\mathbf{K}$ matrix here is based on the initial PSF model, before the multi-Gaussian fitting. The posterior of the PSF and AGN parameters is given by

$$P(\boldsymbol{\nu}, \boldsymbol{\xi}|\boldsymbol{d}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\zeta}}) = \frac{P(\boldsymbol{d}|\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\zeta}}, \boldsymbol{\nu}, \boldsymbol{\xi})P(\boldsymbol{\nu}, \boldsymbol{\xi})}{P(\boldsymbol{d}|\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\zeta}})}, \quad (26)$$

where $\boldsymbol{\nu}$ represents the parameters of the Gaussian profiles in equation (24) that yield $\boldsymbol{w}_{[0]}$; $\boldsymbol{\xi}$ are the amplitudes and the positions of the AGN, which are coded in $\mathbf{M}$. The likelihood of equation (26) is

$$P(\boldsymbol{d}|\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\zeta}}, \boldsymbol{\nu}, \boldsymbol{\xi}) = \frac{\exp[-E_{\rm D}(\boldsymbol{d}|\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\zeta}}, \boldsymbol{\nu}, \boldsymbol{\xi})]}{Z_{\rm D}}, \quad (27)$$

where

$$E_{\rm D}(\boldsymbol{d}|\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\zeta}}, \boldsymbol{\nu}, \boldsymbol{\xi}) = \frac{1}{2}(\boldsymbol{d} - \mathbf{K}\hat{\boldsymbol{g}} - \mathbf{K}\hat{\mathbf{L}}\hat{\boldsymbol{s}} - \mathbf{M}\boldsymbol{w}_{[0]})^{\rm T} \cdot \mathbf{C}_{\rm D}^{-1}(\boldsymbol{d} - \mathbf{K}\hat{\boldsymbol{g}} - \mathbf{K}\hat{\mathbf{L}}\hat{\boldsymbol{s}} - \mathbf{M}\boldsymbol{w}_{[0]}), \quad (28)$$

and $Z_{\rm D} = (2\pi)^{N_{\rm d}/2}(\det\, \mathbf{C}_{\rm D})^{1/2}$. We denote $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{\xi}}$ as the maximum likelihood parameters [that maximizes equation (26)] from which we can obtain the optimal AGN light on the image, given the optimized source and lens-mass models from the previous sections.

4.3 Pixelated fine structure of AGN light

In this section, we introduce the inner loop which aims at extracting the fine structure, $\delta\boldsymbol{w}_{[0]}$, in equation (22) by using a correction grid.

[6] The different Gaussian components can vary their amplitudes, position angles, and axis ratios.

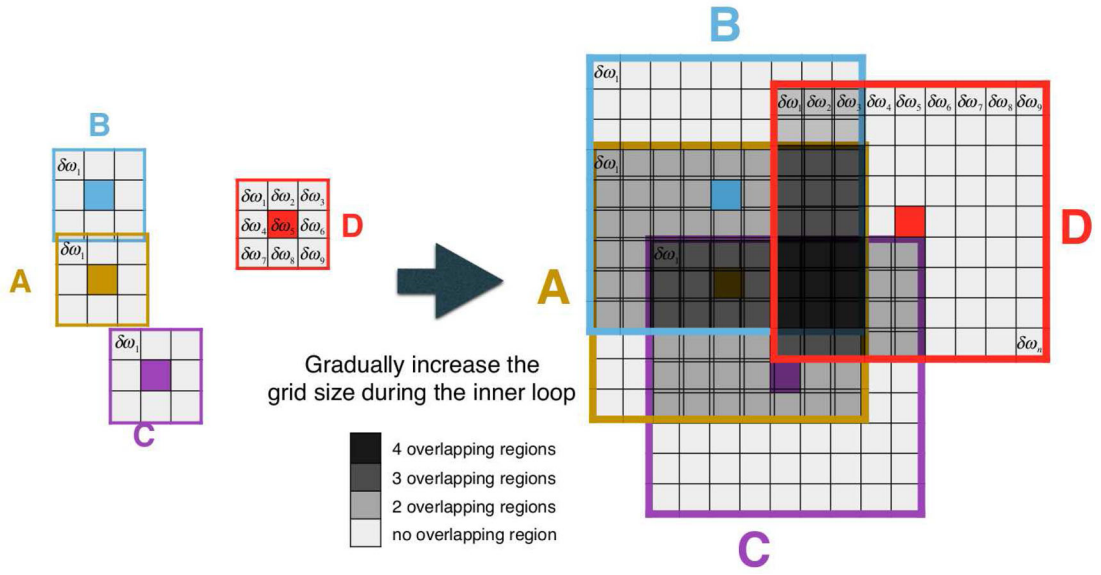

Figure 5. The PSF correction grids of the iterative PSF correction scheme. In the inner loop of the PSF correction scheme (same n but different m), we start with a small correction grid $\delta \boldsymbol{w}$ and increase it sequentially. This accommodates for the larger corrections needed in the central parts of the AGN. Left-hand panel: a small PSF correction grid is placed at each of the four AGN images A, B, C, and D in Fig. 1 via the matrix $\mathbf{M}$, and the values of the PSF correction grid is determined via a linear inversion to reduce the overall image residuals. Since the AGN centroids are typically non-integral pixel values, we linearly interpolate the correction grid on to the image plane. Right-hand panel: the enlarged corrections grids after several iterations of PSF correction, showing overlap between the grids. When the peripheral area of a correction grid overlaps with the central parts of another AGN image (e.g. AGN image C in the lower-right parts of the correction grid of image A), we mask out the centre of the AGN region in order to prevent the correction grids from absorbing the residuals which come from the mismatch of the sharp intensity of AGN centre (see Appendix A for more details).

We show it visually in Fig. 5. The goal of the inner loop is to incorporate most of the fine structures into the PSF model; then in the outer loop, we can use the updated PSF model obtained from the inner loop to remodel all the light components (which require a given PSF model). Since this section is the starting point of the inner loop and outer loop, we obtain $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\zeta}}$, $\hat{\boldsymbol{s}}$, $\hat{\boldsymbol{\nu}}$, and $\hat{\boldsymbol{\xi}}$ by optimizing equations (10), (17), and (26), which are actually the zeroth outer loop iteration and the zeroth inner loop iteration, which we denote by $\hat{\boldsymbol{\eta}}_{[0]}$, $\hat{\boldsymbol{\zeta}}_{[0]}$, $\hat{\boldsymbol{s}}_{[0]}$, $\hat{\boldsymbol{\nu}}_{[0,0]}$, and $\hat{\boldsymbol{\xi}}_{[0,0]}$.

4.3.1 *PSF correction for each iteration (inner loop: step 3)*

In general, given $\hat{\boldsymbol{\eta}}_{[n]}$, $\hat{\boldsymbol{\zeta}}_{[n]}$, $\hat{\boldsymbol{s}}_{[n]}$, $\hat{\boldsymbol{\nu}}_{[n,m]}$[7], and $\hat{\boldsymbol{\xi}}_{[n,m]}$, where m is the iteration number of the inner loop and n is the iteration number of the outer loop, we can write out the equation as

$$\boldsymbol{d}^{\mathrm{P}} = \mathbf{K}_{[n]}\hat{\boldsymbol{g}}_{[n]} + \mathbf{K}_{[n]}\hat{\mathbf{L}}_{[n]}\hat{\boldsymbol{s}}_{[n]} + \hat{\mathbf{M}}_{[n,m]}(\hat{\boldsymbol{w}}_{[n,m]} + \mathbf{T}_{[n,m]}\delta\boldsymbol{w}_{[n,m]}) \equiv \boldsymbol{d}^{\mathrm{P}}_{\mathrm{correction}}, \tag{29}$$

where

$$\hat{\boldsymbol{w}}_{[n,m]} = \begin{cases} \boldsymbol{w}_{[n,m]}(\hat{\boldsymbol{\nu}}_{[n,m]}) & \text{if } n = m = 0 \\ \boldsymbol{w}_{[n,m]} & \text{otherwise} \end{cases}$$

$\hat{\boldsymbol{g}}_{[n]} = \boldsymbol{g}(\hat{\boldsymbol{\eta}}_{[n]})$, $\hat{\mathbf{L}}_{[n]} = \mathbf{L}(\hat{\boldsymbol{\zeta}}_{[n]})$, $\hat{\mathbf{M}}_{[n,m]} = \mathbf{M}(\hat{\boldsymbol{\xi}}_{[n,m]})$, $\mathbf{K}_{[n]}$ is the nth blurring matrix (we explain how to get the nth blurring matrix in Section 4.4.1), $\mathbf{T}_{[n,m]}$ is a matrix which makes $\delta\boldsymbol{w}_{[n,m]}$ the same length as $\hat{\boldsymbol{w}}_{[n,m]}$ by padding with zeros (we show it visually in Appendix B), and $\delta\boldsymbol{w}_{[n,m]}$ is the fine structure we want to obtain by the end of this section.

[7] $\hat{\boldsymbol{\nu}}_{[n,m]}$ is only present when $n = m = 0$, which corresponds to parameters of the Gaussian profiles in equation (24).

The posterior of $\delta\boldsymbol{w}_{[n,m]}$ is

$$P(\delta\boldsymbol{w}_{[n,m]}|\boldsymbol{d}, \hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \hat{\boldsymbol{\nu}}_{[n,m]}, \hat{\boldsymbol{\xi}}_{[n,m]}, \lambda_{\delta w,[n,m]}, R) = \frac{P(\boldsymbol{d}|\delta\boldsymbol{w}_{[n,m]}, \hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \hat{\boldsymbol{\nu}}_{[n,m]}, \hat{\boldsymbol{\xi}}_{[n,m]})}{P(\boldsymbol{d}|\lambda_{\delta w,[n,m]}, \hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \hat{\boldsymbol{\nu}}_{[n,m]}, \hat{\boldsymbol{\xi}}_{[n,m]}, R)} \cdot P(\delta\boldsymbol{w}_{[n,m]}|\lambda_{\delta w,[n,m]}, R), \tag{30}$$

where $P(\delta\boldsymbol{w}_{[n,m]}|\lambda_{\delta w,[n,m]}, R)$ is the prior on $\delta\boldsymbol{w}_{[n,m]}$ given $\{\lambda_{\delta w,[n,m]}, R\}$ with R denoting a particular form of 'regularization' on $\delta\boldsymbol{w}_{[n,m]}$ and $\lambda_{\delta w,[n,m]}$ characterizing the strength of the regularization. We can write the likelihood in equation (30) as

$$P(\boldsymbol{d}|\delta\boldsymbol{w}_{[n,m]}, \hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \hat{\boldsymbol{\nu}}_{[n,m]}, \hat{\boldsymbol{\xi}}_{[n,m]}) = \frac{\exp[-E_{\mathrm{D,mAc},[n,m]}(\boldsymbol{d}|\delta\boldsymbol{w}_{[n,m]}, \hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \hat{\boldsymbol{\nu}}_{[n,m]}, \hat{\boldsymbol{\xi}}_{[n,m]})]}{Z_{\mathrm{D,mAc}}}, \tag{31}$$

where 'mAc' stands for maskAGNcenter,

$$E_{\mathrm{D,mAc},[n,m]}(\boldsymbol{d}|\delta\boldsymbol{w}_{[n,m]}, \hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \hat{\boldsymbol{\nu}}_{[n,m]}, \hat{\boldsymbol{\xi}}_{[n,m]}) = \frac{1}{2}(\boldsymbol{d} - \boldsymbol{d}^{\mathrm{P}}_{\mathrm{correction}})^{\mathrm{T}}\mathbf{C}^{-1}_{\mathrm{D,mAc}}(\boldsymbol{d} - \boldsymbol{d}^{\mathrm{P}}_{\mathrm{correction}}), \tag{32}$$

and $Z_{\mathrm{D,mAc}} = (2\pi)^{N_{\mathrm{d}}/2}(\det \mathbf{C}_{\mathrm{D,mAc}})^{1/2}$ is the normalization for the probability. We discuss the mAc (maskAGNcenter) regions in Appendix A.

The prior/regularization we impose in equation (30) on the correction grid (fine structure of PSF) is to prevent the correction grid from absorbing the noise in the observed image. We express the prior in the following form

$$P(\delta\boldsymbol{w}_{[n,m]}|\lambda_{\delta w,[n,m]}, R) = \frac{\exp(-\lambda_{\delta w,[m]}E_{\delta w,[n,m]}(\delta\boldsymbol{w}_{[n,m]}|R))}{Z_{\delta w,[n,m]}(\lambda_{\delta w,[n,m]})}, \tag{33}$$

where $\lambda_{\delta w,[n,m]}$ is the regularization constant of correction, $Z_{\delta w,[n,m]}(\lambda_{\delta w,[n,m]}) = \int d^{N_{\delta w,[n,m]}}\delta \boldsymbol{w}_{[n,m]} \exp(-\lambda_{\delta w,[n,m]} E_{\delta w,[n,m]})$ is the normalization of the prior probability distribution (note that the optimal $\lambda_{\delta w,[n,m]}$ is not determined yet), and $N_{\delta w,[n,m]}$ is the number of pixels of the correction grid. We use the curvature form for the function $E_{\delta w,[n,m]}$, which is discussed in Suyu et al. (2006).

Again, it is easy to understand that we want to maximize equation (30). We obtain the most probable solution

$$\delta \boldsymbol{w}_{[n,m]} = (\boldsymbol{F} + \lambda_{\delta w,[n,m]}\boldsymbol{H})^{-1}(\mathbf{M}_{[n,m]}\mathbf{T}_{[n,m]})^{\mathrm{T}}\mathbf{C}^{-1}_{\mathrm{D,mAc}}\boldsymbol{u}, \tag{34}$$

where

$$\begin{aligned}\boldsymbol{F} &= \nabla\nabla E_{\mathrm{D,mAc},[n,m]} \\ &= \mathbf{T}^{\mathrm{T}}_{[n,m]}(\mathbf{M}^{\mathrm{T}}_{[n,m]}\mathbf{C}^{-1}_{\mathrm{D,mAc}}\mathbf{M}_{[n,m]})\mathbf{T}_{[n,m]},\end{aligned} \tag{35}$$

$$\boldsymbol{H} = \nabla\nabla E_{\delta w,[n,m]}, \tag{36}$$

$$\boldsymbol{u} = \boldsymbol{d} - \mathbf{K}_{[n]}\hat{\boldsymbol{g}}_{[n]} - \mathbf{K}_{[n]}\hat{\mathbf{L}}_{[n]}\boldsymbol{s}_{[n]} - \hat{\mathbf{M}}_{[n,m]}\hat{\boldsymbol{w}}_{[n,m]}, \tag{37}$$

and

$$\nabla \equiv \frac{\partial}{\partial \delta \boldsymbol{w}_{[n,m]}}. \tag{38}$$

Now, we go back to find the optimal regularization constant; that is, we want to maximize

$$P(\lambda_{\delta w,[n,m]}|\boldsymbol{d}, R) = \frac{P(\boldsymbol{d}|R, \lambda_{\delta w[,[n,m]})P(\lambda_{\delta w,[n,m]})}{P(\boldsymbol{d}|R)} \tag{39}$$

using Bayes' rule. If we assume a flat prior in log $\lambda_{\delta w,[n,m]}$, we want to maximize $P(\boldsymbol{d}|R, \lambda_{\delta w,[n,m]})$, which is the evidence in equation (30). Following the results from Suyu et al. (2006), we get

$$\begin{aligned}&2\hat{\lambda}_{\delta w,[n,m]}E_{\delta w,[n,m]}(\delta \boldsymbol{w}_{[n,m]}) \\ &= N_{\delta w,[n,m]} - \hat{\lambda}_{\delta w,[n,m]}\mathrm{Tr}[(\boldsymbol{F} + \hat{\lambda}_{\delta w,[n,m]}\boldsymbol{H})^{-1}\boldsymbol{H}],\end{aligned} \tag{40}$$

where Tr denotes the trace and $\hat{\lambda}_{\delta w,[n,m]}$ is the optimal regularization constant. If we set $m = 0$ (zeroth iteration of the fine structure), we obtain $\delta \boldsymbol{w}_{[n,0]}$. Due to the sharp intensity of the AGN centre, the residuals there are much stronger than the peripheral area. If we directly extract the full correction grid, the regularization intends to under-regularize on the peripheral area and over-regularize on the centre. To avoid this problem, at first, we extract the correction only around the AGN centre; that is, we start from small $N_{\delta w,[n,m]}$ (half-light radius or smaller) and increase it gradually (around 1.2 times previous size each time) as we obtain $\delta \boldsymbol{w}_{[n,m]}$. We show the idea in Fig. 5 (note that the indices on $\delta \boldsymbol{w}$ in the figure are labelling the pixels, rather than the iteration numbers).

Since every iteration of $\delta \boldsymbol{w}_{[n,m]}$ has their own fine structure (correction) uncertainty, according to Suyu et al. (2006), we also take as estimates of the 1σ uncertainty on each pixel value the square root of the corresponding diagonal element of the covariance matrix given by

$$\mathbf{C}_{\delta w,[n,m]} = (\boldsymbol{F} + \hat{\lambda}_{\delta w,[n,m]}\boldsymbol{H})^{-1}. \tag{41}$$

4.3.2 *Add fine structure into global structure (inner loop: step 4)*

We start with the zeroth inner loop iteration, by setting $m = 0$, of the global structure, $\boldsymbol{w}_{[n,0]}$, and fine structure, $\delta \boldsymbol{w}_{[n,0]}$ (which we can obtain by following the previous two sections). We then add the fine structure into the global structure by defining

$$\boldsymbol{w}_{[n,1]} = \boldsymbol{w}_{[n,0]} + \mathbf{T}_{[n,0]}\delta \boldsymbol{w}_{[n,0]}, \tag{42}$$

where $\boldsymbol{w}_{[n,1]}$ is the first iteration in inner loop. More generally, we define the $(m + 1)$th iteration of the PSF as

$$\boldsymbol{w}_{[n,m+1]} = \boldsymbol{w}_{[n,m]} + \mathbf{T}_{[n,m]}\delta \boldsymbol{w}_{[n,m]}. \tag{43}$$

We recalculate the AGN parameters every time after getting a new $\boldsymbol{w}_{[n,m+1]}$, so given the same $\hat{\boldsymbol{\eta}}_{[n]}$ and $\hat{\boldsymbol{\zeta}}_{[n]}$ in equation (29), the posterior of the AGN parameters is given by

$$\begin{aligned}&P(\boldsymbol{\xi}_{[n,m+1]}|\boldsymbol{d}, \hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}) \\ &= \frac{P(\boldsymbol{d}|\hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \boldsymbol{\xi}_{[n,m+1]})P(\boldsymbol{\xi}_{[n,m+1]})}{P(\boldsymbol{d}|\hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]})}.\end{aligned} \tag{44}$$

(Recall that $\boldsymbol{\xi}_{[n,m+1]}$ represents the relative amplitudes and the positions of the AGNs in the nth outer loop iteration, and $(m + 1)$th inner loop iteration.) The likelihood of equation (44) is

$$\begin{aligned}&P(\boldsymbol{d}|\hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \boldsymbol{\xi}_{[n,m+1]}) \\ &= \frac{\exp[-E_{\mathrm{D},[n,m+1]}(\boldsymbol{d}|\hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \boldsymbol{\xi}_{[n,m+1]})]}{Z_{\mathrm{D}}},\end{aligned} \tag{45}$$

where

$$\begin{aligned}&E_{\mathrm{D},[n,m+1]}(\boldsymbol{d}|\hat{\boldsymbol{\eta}}_{[n]}, \hat{\boldsymbol{\zeta}}_{[n]}, \boldsymbol{\xi}_{[n,m+1]}) \\ &= \frac{1}{2}(\boldsymbol{d} - \boldsymbol{\Omega})^{\mathrm{T}}\mathbf{C}^{-1}_{\mathrm{D}}(\boldsymbol{d} - \boldsymbol{\Omega})\end{aligned} \tag{46}$$

with

$$\boldsymbol{\Omega} = \mathbf{K}_{[n]}\hat{\boldsymbol{g}}_{[n]} + \mathbf{K}_{[n]}\hat{\mathbf{L}}_{[n]}\hat{\boldsymbol{s}}_{[n]} + \mathbf{M}_{[n,m+1]}\boldsymbol{w}_{[n,m+1]}, \tag{47}$$

and Z_{D}, as usual, is $(2\pi)^{N_{\mathrm{d}}/2}(\det\ \mathbf{C}_{\mathrm{D}})^{1/2}$. After maximizing equation (44), we obtain $\hat{\boldsymbol{\xi}}_{[n,m+1]}$. We then replace the $\hat{\boldsymbol{\xi}}_{[n,m]}$ from the previous iteration with the $\hat{\boldsymbol{\xi}}_{[n,m+1]}$ we just obtained, and conduct the next inner loop iteration.

4.3.3 *The criteria to stop the inner loop*

During every inner loop, we gradually increase the size, $N_{\delta w,[n,m]}$, of the correction grid. Then, if (1) there is no residuals outside the correction grid, (2) equation (34) has no intensity, and (3) equation (46) no longer decreases, we stop the inner loop. Assuming we have N_{inner} iterations in the inner loop, we obtain $\boldsymbol{w}_{[n,N_{\mathrm{inner}}]}$ and $\hat{\boldsymbol{\xi}}_{[n,N_{\mathrm{inner}}]}$. We then define

$$\boldsymbol{w}_{[n,N_{\mathrm{inner}}]} \equiv \boldsymbol{w}_{[n+1,0]} \equiv \boldsymbol{w}_{[n+1]} \tag{48}$$

and

$$\hat{\boldsymbol{\xi}}_{[n,N_{\mathrm{inner}}]} \equiv \hat{\boldsymbol{\xi}}_{[n+1,0]} \equiv \hat{\boldsymbol{\xi}}_{[n+1]}. \tag{49}$$

4.4 Lens modelling with updated PSF

The goal of the outer loop in Fig. 2 is to remodel all the light components with the updated PSF; that is, we want to obtain a better lens light model and arc light model with the new blurring matrix, and the underlying fine structure can then be revealed.

4.4.1 *Update the blurring matrix and the image covariance matrix (outer loop: step 6)*

Blurring matrix : after obtaining the last version of the PSF from Section 4.3.3, we update the blurring matrix, **K**, in equation (7). In order to accelerate modelling speed, which highly depends on the size of the PSF for convolution of the extended images, we

choose the central $l_{[n]} \times l_{[n]}$ pixels of the updated PSF grid (that has $N_{\delta w,\,[n,\,N\mathrm{inner}]}$ pixels) as the new PSF to construct $\mathbf{K}_{[n+1]}$ for the spatially extended images.[8] Image covariance matrix : we accumulate the uncertainty of the PSF pixel grid from every inner loop. The accumulated uncertainty is

$$n^2_{\delta w,[n+1],k} = \sum_{m=0}^{N_{\mathrm{inner}}} \sum_i \mathrm{T}_{[n,m],ki} \mathrm{C}_{\delta w,[n,m],ij} \delta_{ij}, \tag{50}$$

where $\mathrm{T}_{[n,\,m],\,ki}$ is the element at k row and i column of $\mathbf{T}_{[n,\,m]}$, $\mathrm{C}_{\delta w,\,[n,\,m],\,ij}$ is the element at i row and j column of $\mathbf{C}_{\delta w,\,[n,\,m]}$, and δ_{ij} is the Kronecker delta. The element of the $(n+1)$th noise vector is defined as

$$n_{[n+1],\mu} = \sqrt{n^2_\mu + \sum_k \hat{\mathrm{M}}_{[n+1],\mu k} n^2_{\delta w,[n+1],k}}, \tag{51}$$

which is characterized by the covariance matrix $\mathbf{C}_{\mathrm{D},\,[n\,+\,1]}$[9]. Note that n_μ is the element of the original data noise vector.

4.4.2 Lens modelling with all light components (outer loop: step 2)

In general, when executing the next iteration of outer loop, we can express equation (7) as

$$\begin{aligned} \boldsymbol{d}^{\mathrm{P}} &= \mathbf{K}_{[n+1]} \boldsymbol{g}_{[n+1]} + \mathbf{K}_{[n+1]} \mathbf{L}_{[n+1]} \boldsymbol{s}_{[n+1]} + \mathbf{M}_{[n+1]} \boldsymbol{w}_{[n+1]} \\ &\equiv \boldsymbol{d}^{\mathrm{P}}_{\mathrm{total}}. \end{aligned} \tag{52}$$

The posterior can be written as

$$\begin{aligned} &P(\boldsymbol{\eta}_{[n+1]}, \boldsymbol{\zeta}_{[n+1]}, \boldsymbol{\xi}_{[n+1]} | \boldsymbol{d}, \Delta \boldsymbol{t}) \\ &\propto P(\boldsymbol{d}, \Delta \boldsymbol{t} | \boldsymbol{\eta}_{[n+1]}, \boldsymbol{\zeta}_{[n+1]}, \boldsymbol{\xi}_{[n+1]}) P(\boldsymbol{\eta}_{[n+1]}, \boldsymbol{\zeta}_{[n+1]}, \boldsymbol{\xi}_{[n+1]}). \end{aligned} \tag{53}$$

The likelihood of the data can be expressed as

$$\begin{aligned} &P(\boldsymbol{d}, \Delta \boldsymbol{t} | \boldsymbol{\eta}_{[n+1]}, \boldsymbol{\zeta}_{[n+1]}, \boldsymbol{\xi}_{[n+1]}) \\ &= \int \mathrm{d}\boldsymbol{s}_{[n+1]}\, \mathrm{P}(\boldsymbol{d}, \Delta \boldsymbol{t} | \boldsymbol{\eta}_{[n+1]}, \boldsymbol{\zeta}_{[n+1]}, \boldsymbol{\xi}_{[n+1]}, \boldsymbol{s}_{[n+1]}) \mathrm{P}(\boldsymbol{s}_{[n+1]}), \end{aligned} \tag{54}$$

where

$$\begin{aligned} &P(\boldsymbol{d}, \Delta \boldsymbol{t} | \boldsymbol{\eta}_{[n+1]}, \boldsymbol{\zeta}_{[n+1]}, \boldsymbol{\xi}_{[n+1]}, \boldsymbol{s}_{[n+1]}) \\ &= \frac{\exp[-E_{\mathrm{D},[n+1]}(\boldsymbol{d} | \boldsymbol{\eta}_{[n+1]}, \boldsymbol{\zeta}_{[n+1]}, \boldsymbol{\xi}_{[n+1]}, \boldsymbol{s}_{n+1})]}{Z_{\mathrm{D},[n+1]}} \\ &\quad \cdot \prod_{i=1}^{N_{\mathrm{AGN}}} \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{AGN},i}} \exp\left(- \frac{|\hat{\boldsymbol{\theta}}_{\mathrm{AGN},i,[n+1]} - \boldsymbol{\theta}^{\mathrm{P}}_{\mathrm{AGN,i,[n+1]}}|^2}{2\sigma^2_{\mathrm{AGN},i}} \right) \\ &\quad \cdot \prod_{i=1} \frac{1}{\sqrt{2\pi}\sigma_{\Delta t,i}} \exp\left[- \frac{(\Delta t_i - \Delta t^{\mathrm{P}}_{i,[n+1]})^2}{2\sigma^2_{\Delta t,i}} \right], \end{aligned} \tag{55}$$

$$\begin{aligned} &E_{\mathrm{D},[n+1]}(\boldsymbol{d} | \boldsymbol{\eta}_{[n+1]}, \boldsymbol{\zeta}_{[n+1]}, \boldsymbol{\xi}_{[n+1]}, \boldsymbol{s}_{[n+1]}) \\ &= \frac{1}{2} (\boldsymbol{d} - \boldsymbol{d}^{\mathrm{P}}_{\mathrm{total}})^{\mathrm{T}} \mathbf{C}^{-1}_{\mathrm{D},[n+1]} (\boldsymbol{d} - \boldsymbol{d}^{\mathrm{P}}_{\mathrm{total}}), \end{aligned} \tag{56}$$

where

$$Z_{\mathrm{D},[n+1]} = (2\pi)^{N_d/2} (\det\, \mathbf{C}_{\mathrm{D},[n+1]})^{1/2} \tag{57}$$

is the normalization for the probability, and $\hat{\boldsymbol{\theta}}_{\mathrm{AGN},i,[n+1]} = \boldsymbol{\theta}_{\mathrm{AGN},i}(\hat{\boldsymbol{\xi}}_{[n+1]})$.

After maximizing equation (53), we obtain $\hat{\boldsymbol{\eta}}_{[n+1]}$, $\hat{\boldsymbol{\zeta}}_{[n+1]}$, and $\hat{\boldsymbol{\xi}}_{[n+1]}$. Then, we replace the $\hat{\boldsymbol{\eta}}_{[n]}$, $\hat{\boldsymbol{\zeta}}_{[n]}$, and $\hat{\boldsymbol{\xi}}_{[n]}$ in Section 4.3 with $\hat{\boldsymbol{\eta}}_{[n+1]}$, $\hat{\boldsymbol{\zeta}}_{[n+1]}$, and $\hat{\boldsymbol{\xi}}_{[n+1]}$ and then execute the next set of inner loop iterations. If we have a total of N outer loop iterations, we obtain the final $\mathbf{K}_{[N]}$ and $\boldsymbol{w}_{[N]}$.

4.4.3 The criteria to stop the outer loop.

We iterate the outer loop until equation (56) does not decrease.[10] We also ensure that the size of the PSF ($l_{[n]} \times l_{[n]}$) for convolution of the lens light and arc light is big enough. Since the AO PSF can have substantial wings that contribute significantly, the size of the PSF in AO image is usually substantially larger than those of *HST* images. We set the size of the PSF ($l_{[n]} \times l_{[n]}$) such that the modelling results remain stable beyond this PSF size.

5 DEMONSTRATION AND BLIND TEST

In this section, we demonstrate the method using two mock data sets that are created with different PSFs, and show that we can recover the input parameters in both mocks by using the strategy in Section 4 together with GLEE. SHS simulates AO images that mimic the strong-lensing system, RXJ 1131−1231, with two foreground lens galaxies and a background source comprised of an AGN and its host galaxy. SHS uses an elliptically symmetric power-law profile to describe the main lens-mass distribution and a pseudo-isothermal elliptic mass profile to describe the mass distribution of the satellite galaxy. The background host galaxy of the AGN is described by a Sérsic profile with additional star-forming regions superposed, and the lens light distribution is based on a composite of two Sérsic light profiles. The simulated lensed images and background sources are shown in the third and second column, respectively, of the first (mock #1) and third (mock #2) rows of Fig. 6. The difference between the two mocks is their PSFs. In mock #1, the PSF is taken to be a star observed with Keck's laser guide star adaptive optics system (LGSAO) that is relatively sharp and with a lot of structures [full width at half-maximum (FWHM) is ∼0.06 arcsec]. In mock #2, the PSF is relatively diffuse and without structures, which is similar to the PSF in the real data (FWHM is ∼0.09 arcsec). We show them in the first column of the first and third rows of Fig. 6. GCFC does a blind test of the PSF reconstruction method on mock #1; that is, GCFC does not know the input value at the beginning, and SHS only reveals the input value when GCFC has completed the analysis of mock #1. Since the input value is the same in mock #2, GCFC models mock #2 by using the same strategy although the mock #2 test is performed after mock #1 and therefore is not blinded.

5.1 Mock #1: a sharp and rich-structured PSF

The mock #1 image has 200 × 200 surface brightness pixels as constraints. The pixel size is 0.04 arcsec. The simulated time delays in

[8] We increase $l_{[n]}$ during the iterative procedure until the size is sufficiently big such that the scientific results remain stable.

[9] The purpose of updating the image covariance matrix is to speed up the modelling to the final answer since the correction uncertainty that we add into the image covariance matrix is around AGN; that is, we weight the arc region more. However, in the end, if there is no 'correction', equation (50) is close to zero.

[10] $\frac{E_{\mathrm{D},[n]} - E_{\mathrm{D},[n+1]}}{E_{\mathrm{D},[n]}} < 0.2$ per cent.

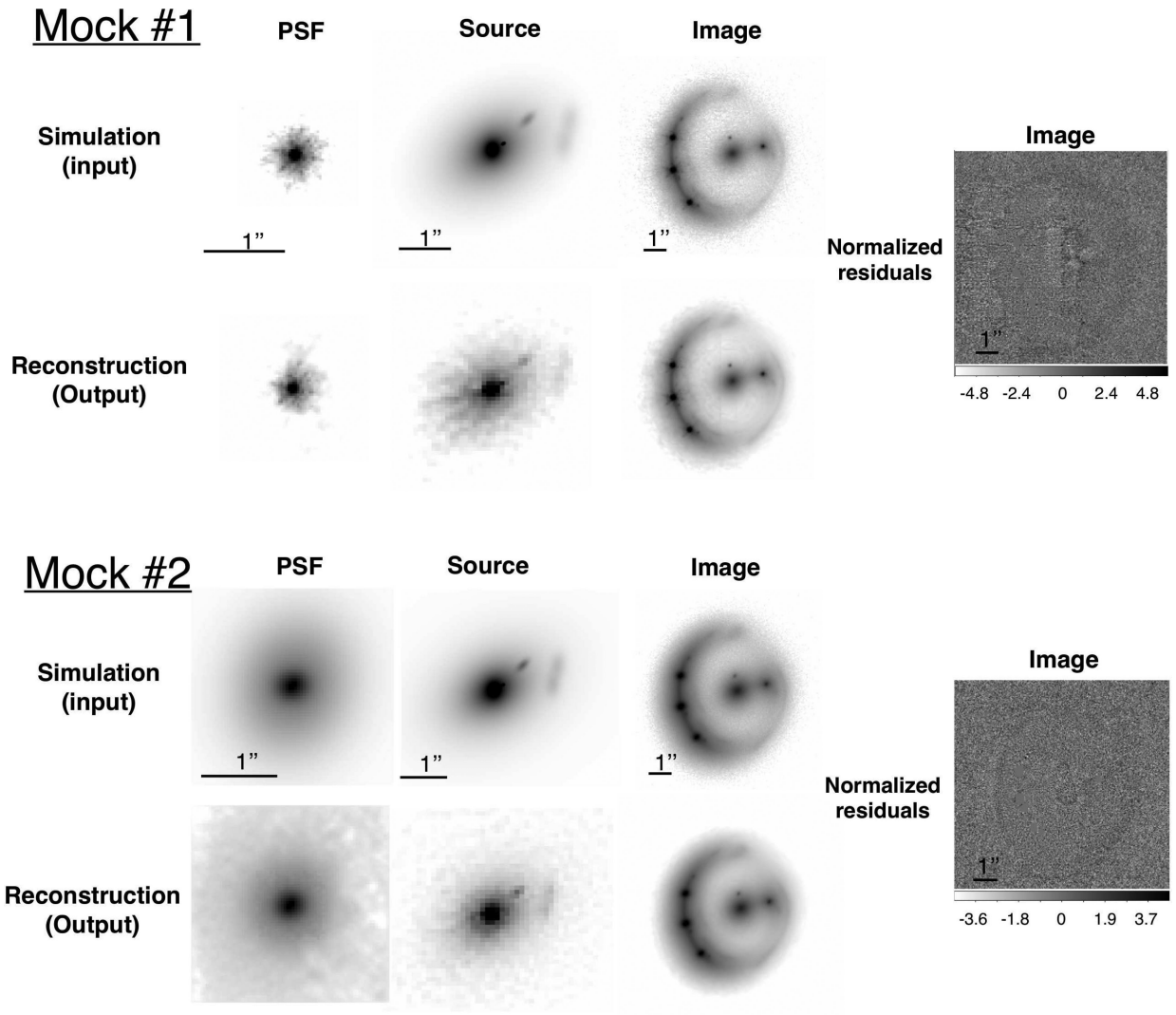

Figure 6. The simulation (input), reconstruction (output), and normalized residuals of mock #1 and mock #2. The left column shows the input/output PSF, the middle-left column shows the input/output sources (host galaxy of the AGN), the middle-right column shows the input/output images, and the right column shows the normalized image residuals (in units of the estimated pixel uncertainties). Our PSF reconstruction method is able to reproduce both the global and fine structures in the PSF, yielding successful reconstructions of the background source intensity and the lensed images. Both reduced χ^2 are $\sim$1.

days relative to image B are: $\Delta t_{\rm AB} = 1.5 \pm 1.5$, $\Delta t_{\rm CB} = -0.5 \pm 1.5$, $\Delta t_{\rm DB} = 90.5 \pm 1.5$.

We follow the procedure described in Section 4 and Fig. 2. The reconstructions are shown in the second row of Fig. 6. To demonstrate the iterative process visually, we show the process in Fig. 7. The first column shows each PSF correction grid in different iteration, the second column shows the cumulative PSF correction from iteration 1 to iteration 18, the third column is the PSF model at each iteration, and the right-most column shows the best-fitting residuals with current PSF model. It is obvious that we get better and better normalized residuals as the iterative PSF corrections proceed. We follow Section 4.3.3 and increase gradually the size of the PSF; the size of the final PSF is 85 $\times$ 85 (which corresponds to 3.4 arcesc $\times$ 3.4 arcesc). However, since the PSF is very sharp in mock #1, the PSF size with 19 $\times$ 19 (which corresponds to 0.76arcesc $\times$ 0.76arcesc) for the blurring matrix is enough. While 19 $\times$ 19 is sufficient for the extended source/lens light, it is not for the AGNs; 85 $\times$ 85 is needed for describing the AGNs.

We try a series of source resolutions from coarse to fine, and the parameter constraints stabilize starting at $\sim$52 $\times$ 52 source pixels, corresponding to source pixel size of $\sim$0.045 arcsec. In order to quantify the systematic uncertainty, we consider the following set of source resolutions: 52 $\times$ 52, 54 $\times$ 54, 56 $\times$ 56, 58 $\times$ 58, 60 $\times$ 60, and 62 $\times$ 62. We weight each choice of the source resolution equally,[11] and combine the Markov chains together. The time delays are also reproduced by the model: for the various source resolutions, the total χ^2 is $\sim$3 for the three delays. We demonstrate the important parameters for cosmography in the upper panel of Fig. 8 (time-delay distance, external shear, radial slope of the main lens galaxy, *Einstein* radii of the main galaxy and its satellite, total *Einstein* radius). The white dots represent the input values. The results show that we can recover the important parameters for cosmography.

[11] We weight the chains by the same weight because the source evidence are similar, and the lens parametrizations are the same.

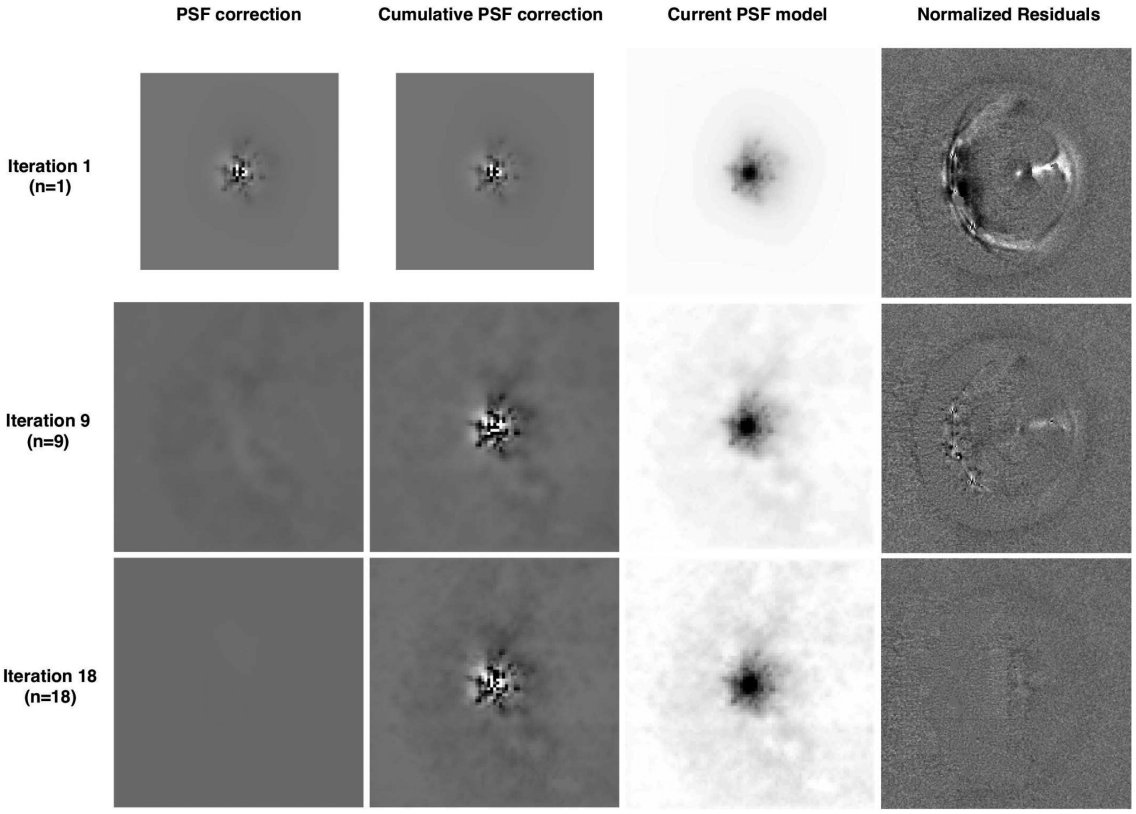

Figure 7. We demonstrate the iterative reconstruction process. From the left to the right, we show the PSF correction, cumulative PSF correction, current PSF model, and normalized residuals after using the current PSF model at iteration 1, 9, and 18. Since we sequentially increase the PSF correction grid as we iterate, the size of the PSF correction grid at iteration 1 is smaller than that of other iterations.

There is a strong degeneracy between the *Einstein* radii of the main galaxy and the satellite galaxy, as expected since these two galaxies are both located within the arcs. However, the effect on time-delay distance due to the presence of the satellite is less than 1 per cent (Suyu et al. 2013). Despite the degeneracy, we can recover the total *Einstein* radius within 1σ, where the total *Einstein* radius, $\theta_{\rm E,\,tot}$, is defined by

$$\frac{\int_0^{\theta_{\rm E,tot}}\int_0^{2\pi}\kappa_{\rm tot}(\theta,\varphi){\rm d}\varphi{\rm d}\theta}{\pi\theta^2_{\rm E,tot}}=1, \tag{58}$$

$\kappa_{\rm tot}$ is the total projected mass density including the main galaxy and its satellite, and φ is the polar angle on the image plane. The total *Einstein* radius in here is only a circular approximation for the elliptical galaxy plus its satellite.

5.2 Mock #2: a diffuse and smooth PSF

The mock #2 image has 300 × 300 surface brightness pixels as constraints (the larger dimensions of the image are necessary for modelling the diffuse PSF). The pixel size and time delays are the same as in mock #1. The size of the final PSF is 127 × 127 (which corresponds to 5.08 arcsec × 5.08 arcsec). Since the PSF is very diffuse in mock #2, the PSF size with 59 × 59 (which corresponds to 2.36 arcsec × 2.36 arcsec) for the blurring matrix is needed to convolve the spatially extended images. We show the reconstruction in the fourth row of Fig. 6.

We also try a series of source resolutions from coarse to fine, and the parameter constraints stabilize starting at ∼59 × 59. To quantify systematic uncertainties due to source resolution, we consider the following set of source resolutions: 59 × 59, 60 × 60, 61 × 61, 62 × 62, and 63 × 63. We also weight each source resolution equally, and combine the Markov chains together. We show the constraints on the same important parameters as mock #1 for cosmography in the lower panel of Fig. 8. The white dots represent the input values. The results show that we can recover the important parameters for cosmography. Again, although we cannot recover the individual *Einstein* radius due to the strong degeneracy between these two *Einstein* radii, we can still recover the total *Einstein* radius.

We use the source-intensity-weighted regularization in the source reconstruction to prevent the source from fitting to the noise. The noise-overfitting problem is due to the fact that the outer region of the source plane is under-regularized. We do two tests which show its negligible impact on cosmographic inference: (1) We test it by changing the image covariance, $\mathbf{C}_{\rm D}$, such that the uncertainties corresponding to low surface brightness areas are boosted (which is a similar effect as allowing the source to be more regularized at low surface brightness regions). The results show that the relative posteriors of lens/cosmological parameters are insensitive to such changes of $\mathbf{C}_{\rm D}$ (hence the source regularization); (2) We impose the source-intensity-weighted regularization on the source plane, which can regularize more on the low surface brightness area on the source plane (see e.g. Tagore & Keeton 2014, for another type of regularization based on analytic profile). Specifically, we obtain the first version of the source intensity distribution $\boldsymbol{s}_{\rm f}$ on a grid of pixels following the method of Suyu et al. (2006) with a constant regularization for all source pixels. We then repeat the source reconstruction but with the regularization constant λ scaled inversely

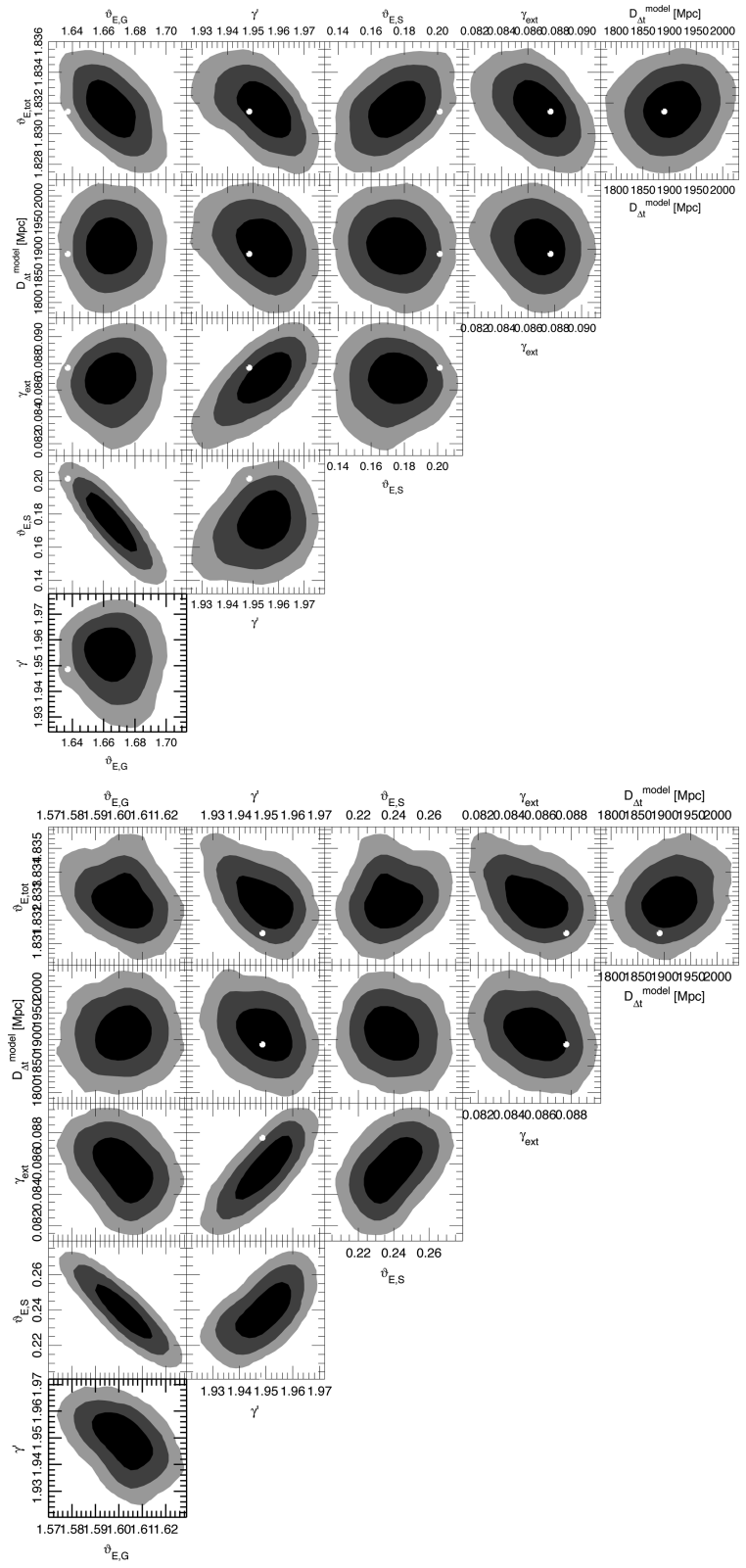

Figure 8. Upper panel: the posterior probability distribution of the key lens model parameters for mock #1. We use the PSF size, 19 × 19, for convolution of the spatially extended images of the AGN host galaxy. We combine the different source resolutions: 52 × 52, 54 × 54, 56 × 56, 58 × 58, 60 × 60, and 62 × 62, and weight each chain equally. The contours/shades mark the 68.3 per cent, 95.4 per cent, and 99.7 per cent credible regions. The white dots are the input values. Lower panel: the posterior probability distribution of the key lens model parameters for mock #2. We use the PSF size, 59 × 59, for convolution of the spatially extended images of the AGN host galaxy. We combine the different source resolutions: 59 × 59, 60 × 60, 61 × 61, 62 × 62, and 63 × 63 and weight each chain equally. We can recover the key lens parameters for cosmography such as the modelled time-delay distance, total *Einstein* radius, and external shear, despite the strong degeneracy between the *Einstein* radii of the main and satellite galaxies (which consequently we do not recover in mock #2).

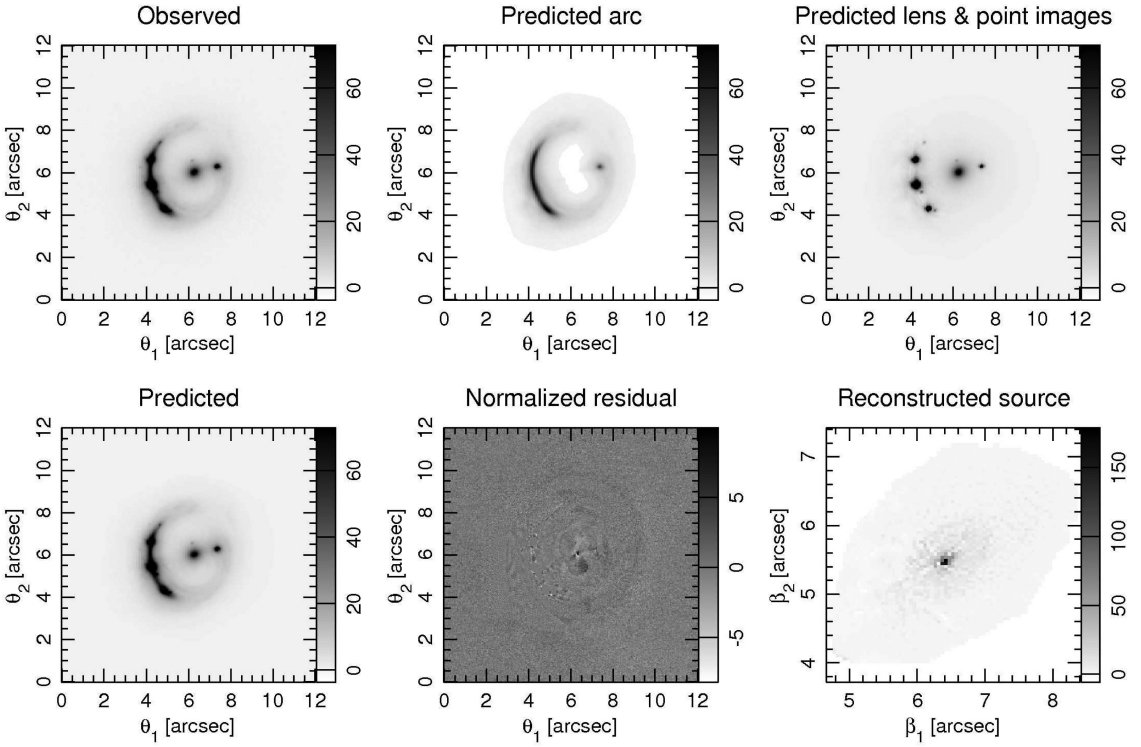

Figure 9. RXJ 1131−1231 AO image reconstruction of the most probable model with a source grid of 79 × 79 pixels and 69 × 69 PSF for convolution of spatially extended images. Top left: RXJ 1131−1231 AO image. Top middle: predicted lensed image of the background AGN host galaxy. Top right: predicted light of the lensed AGNs, the bright compact region: lensed images of a bright compact region in the AGN host galaxy, and the lens galaxies. Bottom left: predicted image from all components, which is a sum of the top-middle and top-right panels. Bottom middle: image residual, normalized by the estimated 1σ uncertainty of each pixel. Bottom right: the reconstructed host galaxy of the AGN in the source plane.

proportional to s_f^4, allowing high/low source intensity regions to be less/more regularized. The relative posteriors between the different MCMC samples in the chains are the same between the uniform and source-intensity-weighted source regularizations. Furthermore, even with different source reconstruction methods, the *Einstein* radius, which also plays an important role in cosmographic inference, is still robust.

6 REAL DATA MODELLING

We apply our newly developed PSF reconstruction method to the real AO imaging shown in Section 2, and use the time delays from Tewes et al. (2013b). For the lens light, we use two Sérsic profiles with common centroids and position angles for the main lens galaxy, and use one circular Sérsic profile for the small satellite (whereas in the mock data in Section 5, we describe the light of the satellite as a point source with PSF, $\boldsymbol{w}$). We find that, in this AO image, four concentric Gaussian profiles provide a good description of the initial global structure of the PSF,[12] which is the procedure we discussed in Section 4.2.3. For modelling the main lens mass, we use an elliptical symmetric distribution with power-law profile and an external shear which are described in Section 4.2.2; for modelling the mass distribution of the satellite, we use a pseudo-isothermal mass distribution.

After we increase the PSF grid during the iterative reconstruction scheme, the final PSF size is 127 × 127 (which corresponds to 5.08 arcsec × 5.08 arcsec). We try a series of source resolutions from coarse to fine and a series of PSF sizes for the blurring matrix from small to large. The parameter constraints stabilize starting at ∼71 × 71 for the source resolution and at ∼59 × 59 for the PSF size for the blurring matrix, corresponding to a source pixel size of ∼0.05 arcsec and a PSF size of 2.36 arcsec × 2.36 arcsec. Note, again, that while a PSF cut-out of 59 × 59 is sufficient for the extended source, the AGNs require a larger PSF grid of 127 × 127. We show the reconstructions of AO imaging in Fig. 9[13] and the reconstructed PSF in Fig. 10. To quantify the systematic uncertainty, we show in Fig. 11 the parameter constraints of different sizes of the source grid, 71 × 71, 73 × 73, 75 × 75, 77 × 77, and 79 × 79, with the PSF size, 59 × 59, for the blurring matrix. After combining all the chains with different source resolutions, we overlap the contours from the 59 × 59 PSF with the contours from the 69 × 69 PSF (for the blurring matrix) in Fig. 12; the results agree with each other within 1σ uncertainty.

Since the PSF in RXJ 1131−1231 AO imaging is similar to the PSF of mock #2, the results from Fig. 8 provide a valuable reference. Thus, note that the *Einstein* radii of the main galaxy and the satellite galaxy inferred from the Keck AO image are also degenerate with each other, as we saw in the case of mock #2.

By using the same time-delay measurements from Tewes et al. (2013b) as in Suyu et al. (2013), we compare the results of

[12] Due to unknown PSF, we do not have prior information on PSF. Thus, we test multiple concentric Gaussian profiles to fit the AGN. However, we find that the initial PSF model does not affect the final results which is shown in Section 5, because the iterative method will correct it in the end.

[13] We use the source-intensity-weighted regularization in the source reconstruction.

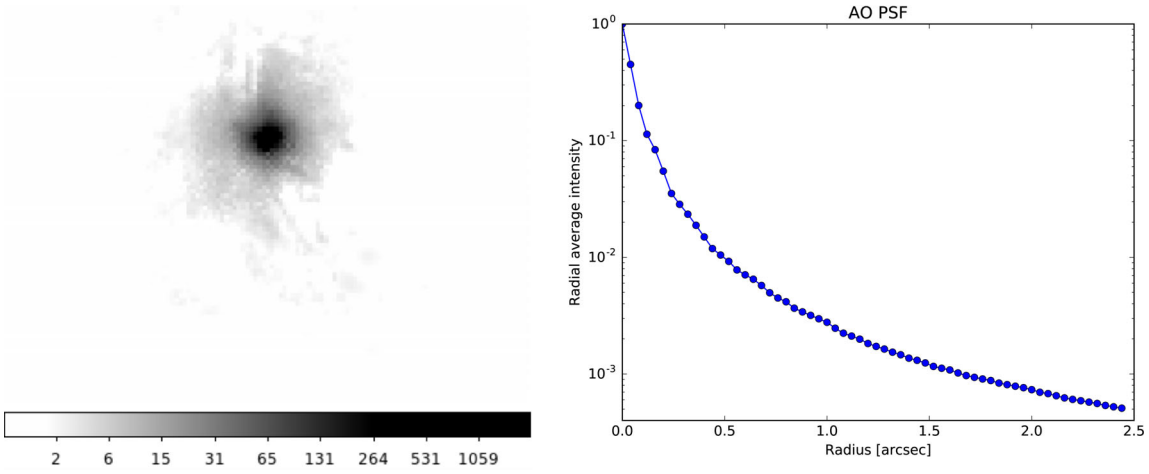

Figure 10. The left-hand panel is the reconstructed AO PSF. The right-hand panel is the radial average intensity of the PSF, which shows the core plus its wings.

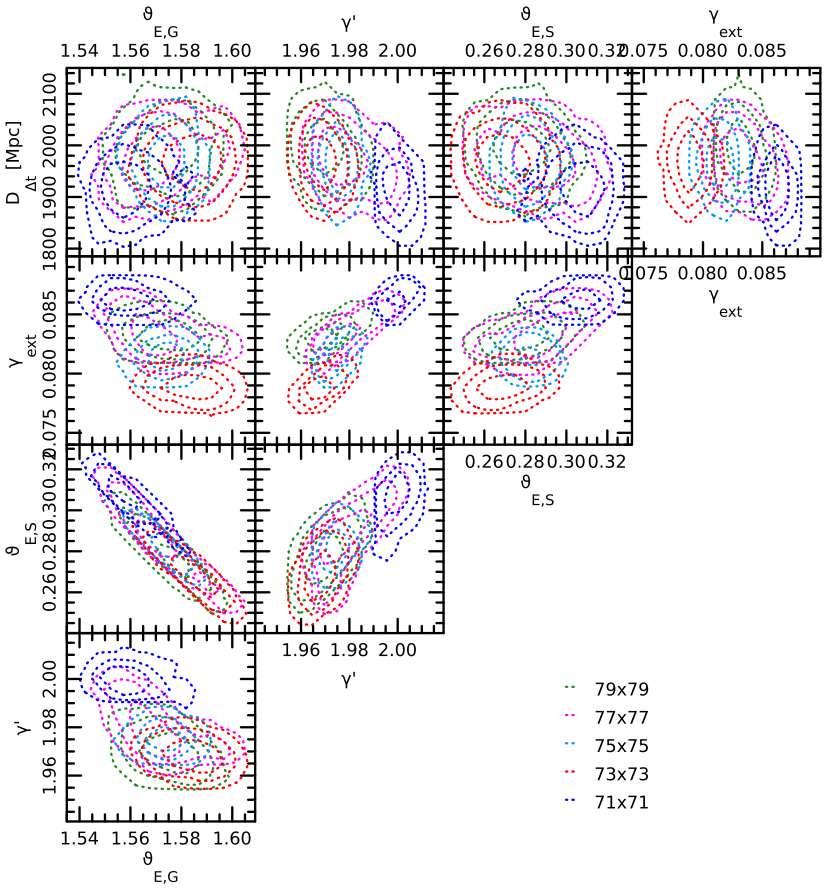

Figure 11. Posterior of the key lens model parameters for RXJ 1131−1231 and the time delays. We use the PSF size, 59 × 59, for convolution of the spatially extended lens and arcs. We show the constraints from Markov chains of different source resolutions: 71 × 71, 73 × 73, 75 × 75, 77 × 77, and 79 × 79. The contours mark the 68.3 per cent, 95.4 per cent, and 99.7 per cent credible regions for each source resolution. The spread in the constraints from different chains allow us to quantify the systematic uncertainty due to the pixelated source resolution.

modelling the AO image with the results of modelling the *HST* image from Suyu et al. (2013).[14] We show the comparison in Fig. 13 and list all the lens model parameters in Table 1. Except for the highly degenerate *Einstein* radius of the main galaxy, other important parameters are overlapping within 1σ uncertainty. Furthermore, the constraint of time-delay distance by using AO imaging

[14] The mass model parametrization is the same as Suyu et al. (2013) except for a slight difference in the definition of θ_E due to ellipticity. In this paper, we compare the θ_E as defined in equation (14). Thus the θ_E shown in this paper is slightly different from that of Suyu et al. (2013).

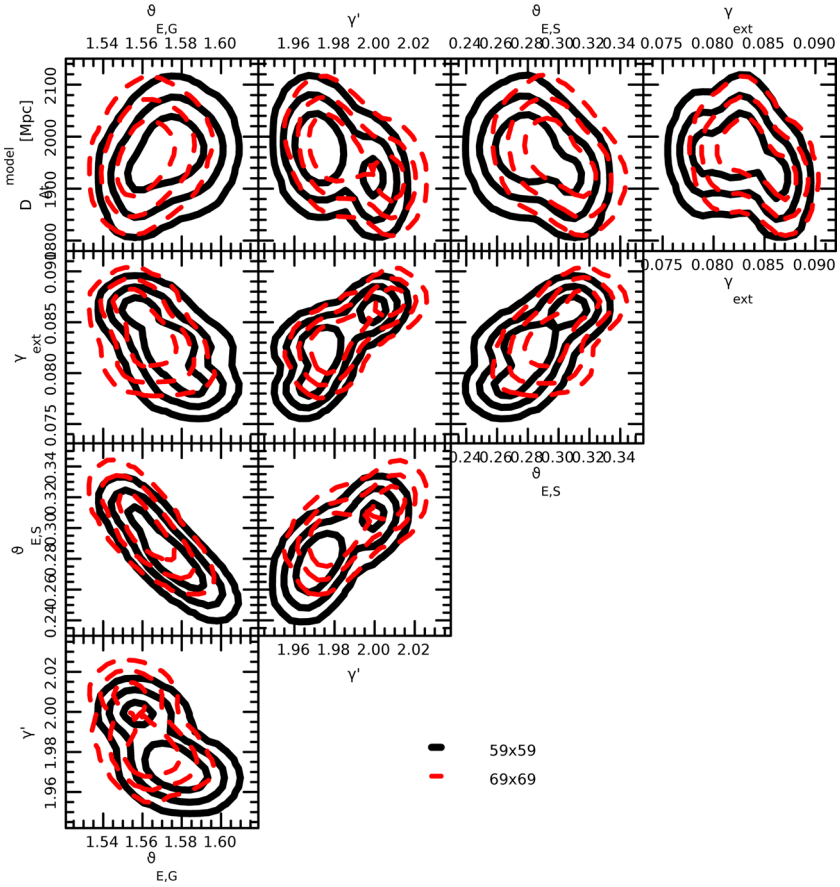

Figure 12. Posterior of the key lens model parameters for RXJ1131 and the time delays. We compare the PSF size, 59 × 59 and 69 × 69, for convolution of the spatially extended lens and arcs. The constraints correspond to the combination of Markov chains of different source resolutions (71 × 71, 73 × 73, 75 × 75, 77 × 77, and 79 × 79) in both PSF sizes. The contours mark the 68.3 per cent, 95.4 per cent, and 99.7 per cent credible regions. The constraints of the two PSF sizes are in good agreement, indicating that PSF sizes larger than ∼59 × 59 are sufficient to capture the PSF features for convolving the spatially extended images.

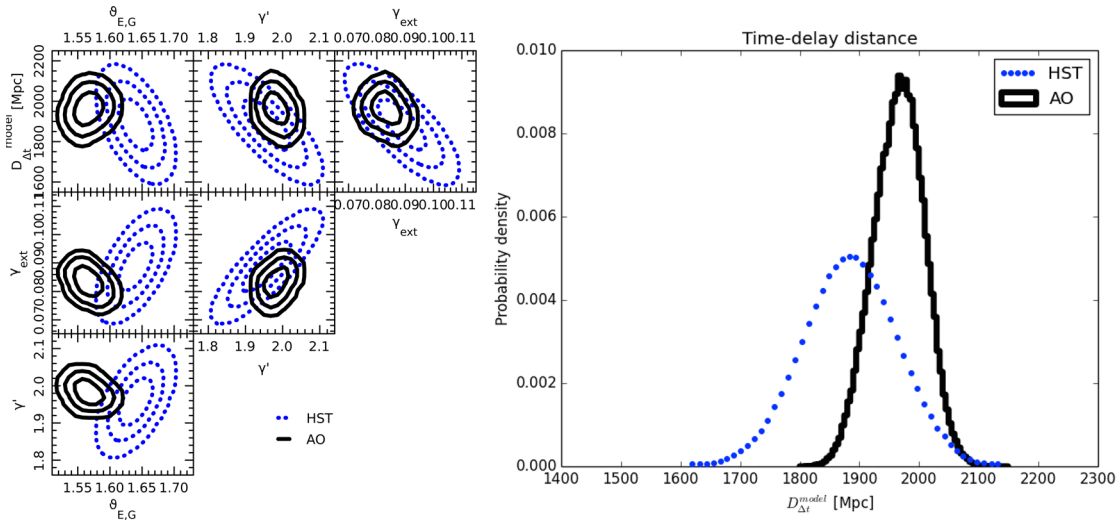

Figure 13. Left-hand panel: comparison of posterior of the key lens model parameters between AO imaging (dashed) and *HST* imaging (shades). The AO constraints are from the combination of both the 59 × 59 and 69 × 69 chains containing the series of source resolutions (e.g. Fig. 12 for 59 × 59). The contours/shades mark the 68.3 per cent, 95.4 per cent, and 99.7 per cent credible regions. The AO constraints are consistent with the *HST* constraints, and are in fact ∼50 per cent tighter on the modelled time-delay distance. Right-hand panel: PDFs for $D_{\Delta t}$, showing the constraints from *HST* image and AO image.

Table 1. Lens model parameter.

Description	Parameter	Marginalized or optimized constraints
Time-delay distance (Mpc)	$D_{\Delta t}^{\rm model}$	1970^{+40}_{-43}
Lens-mass distribution		
Centroid of G in θ_1 (arcsec)	$\theta_{1,\rm G}$[a]	$6.306^{+0.004}_{-0.008}$
Centroid of G in θ_2 (arcsec)	$\theta_{2,\rm G}$	$5.955^{+0.005}_{-0.005}$
Axis ratio of G	$q_{\rm G}$	$0.753^{+0.008}_{-0.007}$
Position angle of G (°)	$\phi_{\rm G}$[b]	$113.4^{+0.4}_{-0.5}$
Einstein radius of G (arcsec)	$\theta_{\rm E,G}$	$1.57^{+0.01}_{-0.01}$
Radial slope of G	γ'	$1.98^{+0.07}_{-0.02}$
Centroid of S in θ_1 (arcsec)	$\theta_{1,\rm S}$	$6.27^{+0.02}_{-0.03}$
Centroid of S in θ_2 (arcsec)	$\theta_{2,\rm S}$	$6.56^{+0.01}_{-0.01}$
Einstein radius of S (arcsec)	$\theta_{\rm E,\,S}$	$0.282^{+0.003}_{-0.003}$
External shear strength	$\gamma_{\rm ext}$	$0.083^{+0.003}_{-0.003}$
External shear angle (°)	$\phi_{\rm ext}$	93^{+1}_{-1}
Lens light as Sérsic profiles		
Centroid of S in θ_1 (arcsec)	$\theta_{1,\rm GL}$	$6.3052^{+0.0002}_{-0.0002}$
Centroid of S in θ_2 (arcsec)	$\theta_{2,\rm GL}$	$6.0660^{+0.0002}_{-0.0002}$
Position angle of G (°)	$\phi_{\rm GL}$	$116.9^{+0.4}_{-0.4}$
Axis ratio of G1	$q_{\rm G}$	$0.912^{+0.004}_{-0.004}$
Amplitude of G1	$I_{\rm s,\,GL1}$[c]	$1.47^{+0.02}_{-0.02}$
Effective radius of G1 (arcsec)	$R_{\rm eff,\,GL1}$	$2.37^{+0.01}_{-0.01}$
Index of G1	$n_{\rm sérsic,GL1}$	$0.63^{+0.01}_{-0.01}$
Axis ratio of G2	$q_{\rm GL2}$	$0.867^{+0.002}_{-0.002}$
Amplitude of G2	$I_{\rm s,\,GL2}$	$18.1^{+0.3}_{-0.3}$
Effective radius of G2 (arcsec)	$R_{\rm eff,\,GL2}$	$0.404^{+0.005}_{-0.005}$
Index of G2	$n_{\rm sérsic,GL2}$	$1.97^{+0.02}_{-0.02}$
Centroid of S in θ_1 (arcsec)	$\theta_{1,\rm SL}$	$6.210^{+0.001}_{-0.001}$
Centroid of S in θ_2 (arcsec)	$\theta_{2,\rm SL}$	$6.605^{+0.001}_{-0.001}$
Axis ratio of S	$q_{\rm SL}$	$\equiv 1$
Amplitude of S	$I_{\rm s,\,SL}$	69^{+6}_{-6}
Effective radius of S (arcsec)	$R_{\rm eff,\,SL}$	$0.027^{+0.001}_{-0.001}$
Index of S	$n_{\rm sérsic,SL}$	$0.42^{+0.04}_{-0.02}$
Lensed AGN light		
Position of image A in θ_1 (arcsec)	$\theta_{1,\rm A}$	4.256
Position of image A in θ_2 (arcsec)	$\theta_{2,\rm A}$	6.652
Amplitude of image A	$a_{\rm A}$	21 880
Position of image B in θ_1 (arcsec)	$\theta_{1,\rm B}$	4.288
Position of image B in θ_2 (arcsec)	$\theta_{2,\rm B}$	4.348
Amplitude of image B	$a_{\rm B}$	38 555
Position of image C in θ_1 (arcsec)	$\theta_{1,\rm C}$	4.871
Position of image C in θ_2 (arcsec)	$\theta_{2,\rm C}$	4.348
Amplitude of image C	$a_{\rm C}$	11 565
Position of image D in θ_1 (arcsec)	$\theta_{1,\rm D}$	7.378
Position of image D in θ_2 (arcsec)	$\theta_{2,\rm D}$	6.340
Amplitude of image D	$a_{\rm D}$	3215

[a]The reference of the position is in Fig. 9.
[b]All the position angles are measured counterclockwise from positive θ_2 (north).
[c]The amplitude is in equation (8).
Note. There are total 39 parameters that are optimized or sampled. The optimal parameters have little effect on the key parameters for cosmology (such as $D_{\Delta t}^{\rm model}$). For the lens light, two Sérsic profiles with common centroid and position angle are used to describe the main lens galaxy G. They are denoted as G1 and G2 above. The source pixel parameters ($\boldsymbol{s}$) are marginalized and are thus not listed.

with 0.09 arcsec resolution is tighter than the constraint of time-delay distance by using *HST* imaging with 0.09 arcsec by around 50 per cent.

For cosmographic measurement from time-delay lenses, we need to break the mass-sheet degeneracy in gravitational lensing (e.g. Falco et al. 1985; Schneider & Sluse 2013, 2014; Xu et al. 2016) that can change the modelled time-delay distance. This would involve considerations of mass profiles, lens stellar kinematics, and external convergence (e.g. Treu & Koopmans 2002; Barnabè et al. 2011; Suyu et al. 2013, 2014) that are beyond the scope of this paper. The focus of this paper is to investigate the feasibility of AO imaging for follow-up. As illustrated in Fig. 13, AO imaging together with our new PSF reconstruction technique (especially of quad lens systems) is a competitive alternative to *HST* imaging for following up time-delay lenses for accurate lens modelling.

7 SUMMARY

In this paper, we develop a new method, namely an iterative PSF correction scheme. This scheme determines the unknown PSF in AO images of gravitational lens systems and thereby overcome the unknown PSF problem to constrain cosmology by modelling the strong-lensing AO imaging with time delays. We elaborate the procedures in Section 4 and draw an overall flow chart in Fig. 2. We test the method on two mock systems, mock #1 (blindly) and mock #2, which are created by using a sharp PSF and diffuse PSF, respectively, and apply this method to the high-resolution AO RXJ 1131−1231 image taken with the Keck telescope as part of the SHARP AO observation. We draw the following conclusions.

(i) We perform a blind test on mock #1, which mimics the appearance of RXJ 1131−1231 but with a sharp and richly structured PSF (based on a star observed with Keck's LGSAO). Afterward, we model the mock #2, which is created by a diffuse PSF that is similar with the PSF in AO RXJ 1131−1231 image, using the same strategy. The results show that the more diffuse PSF the AO imaging has, the larger the PSF is needed for representing the AGN; similarly, the larger the PSF for representing the AGNs, the larger the PSF is needed for convolution of the spatially extended lens and arcs. By performing MCMC sampling, we can recover the important parameters for cosmography (time-delay distance, external shear, slope, and total *Einstein* radius of the main galaxy plus its satellite). Although we cannot recover the individual *Einstein* radius, the effect on time-delay distance due to the presence of the satellite is less than 1 per cent (Suyu et al. 2013).

(ii) We model the AO RXJ 1131−1231 image by the iterative PSF correction scheme. We compare the results of important parameters with the results from modelling the *HST* imaging in Suyu et al. (2013). Except for the highly degenerate *Einstein* radius of the main galaxy, other important parameters for cosmography agree with each other within 1σ (Fig. 13). Furthermore, the constraint of time-delay distance by using AO imaging with 0.09 arcsec resolution is tighter than the constraint of time-delay distance by using *HST* imaging with 0.09 arcsec by around 50 per cent.

The iterative PSF reconstruction method that we have developed is general and widely applicable to studies that require high-precision PSF reconstruction from multiple nearby point sources in the field (e.g. the search of faint companions of stars in star clusters). For the case of gravitational lens time delays, this method lifts the restriction of using *HST* strong-lensing imaging, and opens a new series of AO imaging data set to study cosmology. From the upcoming surveys, hundreds of new lenses are predicted to be discovered;

this method not only can motivate more telescopes to be equipped with AO technology, but also facilitate the goal to reveal possible new physics by beating down the uncertainty on H_0 to 1 per cent via strong lensing (Suyu et al. 2012a).

ACKNOWLEDGEMENTS

We thank Giuseppe Bono, James Chan, Thomas Lai, Anja von der Linden, Eric Linder, Phil Marshall, and David Spergel for the useful discussions. GCFC and SHS are grateful to Bau-Ching Hsieh for computing support on the SuMIRe computing cluster. GCFC and SHS acknowledge support from the Ministry of Science and Technology in Taiwan via grant MOST-103-2112-M-001-003-MY3. SHS and SV acknowledge support by the Max Planck Society through the Max Planck Research Groups. LVEK is supported in part through an NWO-VICI career grant (project number 639.043.308). CDF and DJL acknowledge support from NSF-AST-0909119 and NSF-AST-1312329. CDF also acknowledges support under programme NSF-AST-1312329. The data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

REFERENCES

Agnello A., Kelly B. C., Treu T., Marshall P. J., 2015, MNRAS, 448, 1446
Auger M. W., Treu T., Brewer B. J., Marshall P. J., 2011, MNRAS, 411, L6
Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., 2011, MNRAS, 415, 2215
Bartelmann M., 2010, Class. Quantum Gravity, 27, 233001
Beckers J. M., 1993, ARA&A, 31, 13
Birrer S., Amara A., Refregier A., 2016, J. Cosmol. Astropart. Phys., 8, 20
Brase J. M., 1998, Wavefront control system for the Keck telescope. Available at: http://www.osti.gov/scitech/servlets/purl/302840
Chan J. H. H., Suyu S. H., Chiueh T., More A., Marshall P. J., Coupon J., Oguri M., Price P., 2015, ApJ, 807, 138
Chantry V., Magain P., 2007, A&A, 470, 467
Coe D., Moustakas L. A., 2009, ApJ, 706, 45
Collett T. E. et al., 2013, MNRAS, 432, 679
Courbin F. et al., 2011, A&A, 536, A53
Dunkley J., Bucher M., Ferreira P. G., Moodley K., Skordis C., 2005, MNRAS, 356, 925
Eulaers E. et al., 2013, A&A, 553, A121
Falco E. E., Gorenstein M. V., Shapiro I. I., 1985, ApJ, 289, L1
Fassnacht C. D., Koopmans L. V. E., Wong K. C., 2011, MNRAS, 410, 2167
Freedman W. L., Madore B. F., Scowcroft V., Burns C., Monson A., Persson S. E., Seibert M., Rigby J., 2012, ApJ, 758, 24
Greene Z. S. et al., 2013, ApJ, 768, 39
Hilbert S., White S. D. M., Hartlap J., Schneider P., 2007, MNRAS, 382, 121
Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, A&A, 499, 31
Hinshaw G. et al., 2013, ApJS, 208, 19
Hobson M. P., Bridle S. L., Lahav O., 2002, MNRAS, 335, 377
Hsueh J.-W., Fassnacht C. D., Vegetti S., McKean J. P., Spingola C., Auger M. W., Koopmans L. V. E., Lagattuta D. J., 2016, preprint (arXiv:1601.01671)
Hu W., 2005, in Wolff S. C., Lauer T. R., eds, ASP Conf. Ser. Vol. 339, Observing Dark Energy. Astron. Soc. Pac., San Francisco, p. 215
Jee I., Komatsu E., Suyu S. H., 2015, J. Cosmol. Astropart. Phys., 033
Jee I., Komatsu E., Suyu S. H., Huterer D., 2016, J. Cosmol. Astropart. Phys., 031
Kochanek C. S., Keeton C. R., McLeod B. A., 2001, ApJ, 547, 50
Koopmans L. V. E., Treu T., Fassnacht C. D., Blandford R. D., Surpi G., 2003, ApJ, 599, 70
Lagattuta D. J., Auger M. W., Fassnacht C. D., 2010, ApJ, 716, L185
Lagattuta D. J., Vegetti S., Fassnacht C. D., Auger M. W., Koopmans L. V. E., McKean J. P., 2012, MNRAS, 424, 2800
Linder E. V., 2004, Phys. Rev. D, 70, 043534
MacKay D. J. C., 1992, Neural Comput., 4, 415
Magain P., Courbin F., Sohy S., 1998, ApJ, 494, 472
Marshall P. J., Hobson M. P., Gull S. F., Bridle S. L., 2002, MNRAS, 335, 1037
Marshall P. J. et al., 2016, MNRAS, 455, 1171
Moffat A. F. J., 1969, A&A, 3, 455
More A. et al., 2016, MNRAS, 455, 1191
Oguri M., Marshall P. J., 2010, MNRAS, 405, 2579
Paraficz D., Hjorth J., 2009, A&A, 507, L49
Perlmutter S. et al., 1999, ApJ, 517, 565
Planck Collaboration XIII, 2015, preprint (arXiv:1502.01589)
Rathna Kumar S. et al., 2013, A&A, 557, A44
Refsdal S., 1964, MNRAS, 128, 307
Riess A. G. et al., 1998, AJ, 116, 1009
Riess A. G. et al., 2009, ApJ, 699, 539
Riess A. G. et al., 2011, ApJ, 730, 119
Rousset G., Fontanella J. C., Kern P., Gigan P., Rigaut F., 1990, A&A, 230, L29
Rusu C. E., Oguri M., Minowa Y., Iye M., More A., Inada N., Oya S., 2014, MNRAS, 444, 2561
Rusu C. E. et al., 2016, MNRAS, 458, 2
Schneider P., Sluse D., 2013, A&A, 559, A37
Schneider P., Sluse D., 2014, A&A, 564, A103
Schneider P., Kochanek C. S., Wambsganss J., 2006, Gravitational Lensing: Strong, Weak and Micro. Springer-Verlag, Berlin
Sérsic J. L., 1968, Atlas de galaxias australes. Observatorio Astronomico de Córdoba, Cordoba
Shapiro I. I., 1964, Phys. Rev. Lett., 13, 789
Sluse D. et al., 2003, A&A, 406, L43
Suyu S. H., Halkola A., 2010, A&A, 524, A94
Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983
Suyu S. H., Marshall P. J., Blandford R. D., Fassnacht C. D., Koopmans L. V. E., McKean J. P., Treu T., 2009, ApJ, 691, 277
Suyu S. H., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, ApJ, 711, 201
Suyu S. H. et al., 2012a, preprint (arXiv:1202.4459)
Suyu S. H. et al., 2012b, ApJ, 750, 10
Suyu S. H. et al., 2013, ApJ, 766, 70
Suyu S. H. et al., 2014, ApJ, 788, L35
Tagore A. S., Keeton C. R., 2014, MNRAS, 445, 694
Tewes M., Courbin F., Meylan G., 2013a, A&A, 553, A120
Tewes M. et al., 2013b, A&A, 556, A22
Treu T., 2010, ARA&A, 48, 87
Treu T., Ellis R. S., 2015, Contemp. Phys., 56, 17
Treu T., Koopmans L. V. E., 2002, MNRAS, 337, L6
Treu T., Koopmans L. V. E., 2004, ApJ, 611, 739
van Dam M. A. et al., 2006, PASP, 118, 310
Vegetti S., Lagattuta D. J., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V. E., 2012, Nature, 481, 341
Vuissoz C. et al., 2007, A&A, 464, 845
Vuissoz C. et al., 2008, A&A, 488, 481
Watson J., 1997, Tip Tilt Correction for Astronomical Telescopes using Adaptive Control. Available at: http://www.osti.gov/scitech/servlets/purl/647009
Wizinowich P. L., Le Mignant D., Stomski P. J., Jr, Acton D. S., Contos A. R., Neyman C. R., 2003, Wizinowich P. L., Bonaccini D., eds, Proc.

SPIE Conf. Ser. Vol. 4839, Adaptive Optical System Technologies II. SPIE, Bellingham, p. 9
Wizinowich P. L. et al., 2006, PASP, 118, 297
Xu D., Sluse D., Schneider P., Springel V., Vogelsberger M., Nelson D., Hernquist L., 2016, MNRAS, 456, 739

APPENDIX A: ARC AND AGN MASK REGIONS

We show the three different mask regions, maskArcAGN (mArcAGN), maskArc (mArc), and maskAGNcenter (mAc) in Fig. A1. For modelling the lens light in Section 4.2.1, we mask out the region which contains significant arc light and AGN light in the left-hand panel. For modelling the arc light in Section 4.2.2, we mask out the region with significant AGN light ($>10\sigma$ level) shown in the middle panel. For extracting the PSF correction, we show the residuals in the right-hand panel (which is the image with the lens light, arc light, and AGN light subtracted). When the size of the correction grid is small such that the correction grids do not overlap other AGN centre, we only need to mask out the area where it comes obviously from the host galaxy of AGN. For instance, if the background AGN has compact bright blobs in its host galaxy, due to the limit of the resolution on the source plane, the predicted arc cannot reconstruct the compact blobs, so there are residuals around these compact blobs on the image plane (shown in the right-hand panel with red arrows). In order to prevent the correction grid from absorbing the light due to the resolution problem and adding non-PSF features into the PSF, we mask them out. When the correction grid is enlarged and covers other AGN centres, we need to mask out both regions (AGN centres and lensed compact blob).

APPENDIX B: $\mathbf{T}_{[n,m]}$ MATRIX

Since $\delta\boldsymbol{w}_{[n,m]}$ has different length in each iteration of inner/outer loop $[n, m]$, we use a matrix $\mathbf{T}_{[n,m]}$ to make $\delta\boldsymbol{w}_{[n,m]}$ the same length as $\boldsymbol{w}_{[n,m]}$ by padding the two-dimensional boundaries of the PSF correction grid with zeros, as illustrated in Fig. A2.

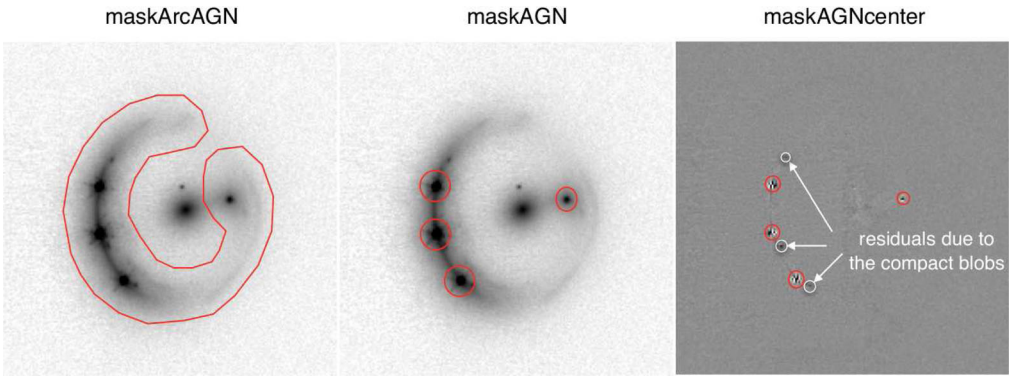

Figure A1. The three different mask regions which are circled in red, and the white arrows indicate the special area which need to be masked out (that is, we boost the uncertainty in that region) while we extract the PSF correction. The left-hand panel shows the maskArcAGN region for fitting the lens light, and the middle panel shows the maskAGN region for fitting the arc light. When obtaining the PSF corrections, the white circles in the right-hand panel need to be masked out when the PSF grid is small. As we increase the PSF grid around each AGN image such that the grid contains other AGN images (shown in the right-hand panel of Fig. 5), we mask out the red circles associated with these other AGN images and also the white circles.

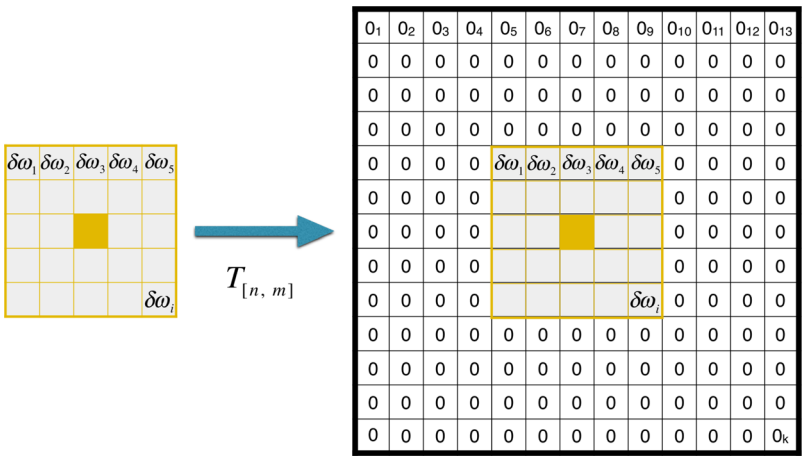

Figure A2. The matrix $\mathbf{T}_{[n,m]}$ for making $\delta\boldsymbol{w}_{[n,m]}$the same length as $\boldsymbol{w}_{[n,m]}$. The indices of $\delta\boldsymbol{w}_i$ and 0_k are for the pixels (rather than the PSF correction iterations). $\mathbf{T}_{[n,m]}$ is a matrix at the n^{th} outer loop and the m^{th} inner loop.

This paper has been typeset from a TEX/LATEX file prepared by the author.