MARS: A New Maximum-entropy-regularized Strong Lensing Mass Reconstruction Method

Sangjun Cha¹ and M. James Jee¹²

¹ Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seoul 03722, Republic of Korea; sang6199@yonsei.ac.kr, mkjee@yonsei.ac.kr
² Department of Physics and Astronomy, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

Received 2022 February 19; revised 2022 April 18; accepted 2022 April 22; published 2022 June 2

Abstract

Free-form strong-lensing (SL) mass reconstructions typically suffer from overfitting, which manifests itself as false-positive small-scale fluctuations. We present a new free-form MAximum-entropy ReconStruction (MARS) method without the assumption that light traces mass (LTM). The MARS algorithm enables us to achieve excellent convergence in source positions (∼0″001), minimize spurious small-scale fluctuations, and provide a quasi-unique solution independently of initial conditions. Our method is tested with the publicly available synthetic SL data FF–SIMS. The comparison with the truth shows that the mass reconstruction quality is on par with those of the best-performing LTM methods published in the literature, which have been demonstrated to outperform existing free-form methods. In terms of the radial mass profile reconstruction, we achieve <1% agreement with the truth for the regions constrained by multiple images. Finally, we apply MARS to A1689 and find that the cluster mass in the SL regime is dominated by the primary halo centered on the brightest cluster galaxy and the weaker secondary halo is also coincident with the bright cluster member ∼160 kpc northeast. Within the SL field, the A1689 radial profile is well described by a Navarro–Frenk–White profile with c200 = 5.53 ± 0.77 and r500 = 538±90/100 kpc, and we find no evidence that A1689 is overconcentrated.

Unified Astronomy Thesaurus concepts: Dark matter distribution (356); Strong gravitational lensing (1643); Galaxy clusters (584)

1. Introduction

Thanks to the advantage of not needing any dynamical assumptions, gravitational lensing is now widely accepted as one of the most powerful methods to measure the mass distribution of galaxy clusters (Bartelmann & Schneider 2001). Depending on the characteristic surface mass density of the lensing system, gravitational lensing is broadly classified into two regimes: strong lensing (SL) and weak lensing (WL).

WL uses measurements of coherent shape distortions of background galaxies to infer mass distributions (e.g., Kaiser & Squires 1993; Hoekstra et al. 2000; Jee et al. 2006; Medezinski et al. 2007; Umetsu et al. 2011; Finner et al. 2017; Kim et al. 2021; Schrabback et al. 2021). Because of the intrinsic shape dispersion of background galaxies, the coherent shape distortions due to lensing are measured by averaging over many source galaxy shapes. Thus, in general, the resolution of the WL mass reconstruction depends on the source number density and is limited.

SL occurs in the regime where the surface mass density of the lens exceeds the so-called critical surface mass density, which is defined by the physical surface mass density and the distance ratios among the observer, lens, and source. SL produces multiple images, which we use for the reconstruction of the responsible mass distribution. Although the mass reconstruction is often limited to the central regions of massive clusters, in general one can achieve a much higher resolution and/or tighter constraint than in WL, in particular near and inside the multiple-image positions. Therefore, SL provides a tremendous opportunity to probe the inner mass profile of the lens in much greater detail (e.g., Broadhurst et al. 2005; Zekser et al. 2006; Julio et al. 2007; Limousin et al. 2007; Zitrin et al. 2009; Coe et al. 2010; Kneib & Natarajan 2011; Lotz et al. 2017; Meneghetti et al. 2017).

Because the number of multiple images for a given lens system is finite, inevitably the two-dimensional mass reconstruction is underdetermined. The traditional method resolves the issue by assuming that light traces mass (LTM). The LTM method is implemented by placing some theoretically motivated analytic halos, such as Navarro–Frenk–White (NFW; Navarro et al. 1996), Pseudo Isothermal Elliptical Mass Distribution (PIEMD; Kassiola & Kovner 1993), etc., on the cluster galaxies and adjusting their properties so that the resulting mass map predicts the observed multiple images. Because the properties of these analytic halos are controlled by their parameters, the LTM approach is also referred to as a parametric mass reconstruction. Typically, the number of free parameters is smaller than the constraints provided by the SL data, and the system is overdetermined. This lack of flexibility often leads to imperfect convergence of multiple-image positions in the source plane.

Although the LTM method is a useful approach as a probe of inner mass profiles of galaxy clusters, it is paramount to investigate SL clusters also with methods that do not rely on the LTM assumption. This alternative approach is essential in validating the LTM hypothesis and may be the only route to probe the truth if the target cluster indeed deviates from our LTM assumptions. This non-LTM approach, also referred to as a free-form or nonparametric method, typically uses a mesh grid to represent the mass distribution.⁴ Compared to the LTM

---

¹ Alternatively and perhaps more precisely, one can also define the SL regime as the region where the sum of the convergence and shear approaches unity.

⁴ Some argue that the term nonparametric here is misleading because in fact the method requires many parameters, which comprise the mesh.
method, the free-form method is flexible because the model is not constrained by the shapes and locations of the halos. The flexibility in general allows the method to converge nearly all legitimate multiple images in the source plane with negligible scatters. However, because of the overwhelming number of free parameters compared to the SL constraints, the system is highly underdetermined, which prevents a unique mass reconstruction and also creates many spurious substructures.

There have been attempts to overcome the drawbacks of the free-form method. One easy solution is to perform the free-form reconstruction numerous times and present the average as the representative mass map. However, although the average mass map may become much smoother than individual results, one cannot guarantee that the average is a legitimate representative solution. One has yet to prove that the sample distribution (from which the average is derived) is a fair representation of the true parent distribution. Also, the average mass map should be able to predict observed SL data (i.e., “delens” multiple-image positions in the image plane to converge in the source plane with negligible scatter and vice versa), which however has not clearly been demonstrated in the literature.

Another approach to prevent overfitting in the non-LTM method is to employ regularization, which imposes “Occam’s razor” on the final solution. The regularization is often implemented by adding an additional function to the likelihood, which prevents the optimization process from creating spurious small-scale fluctuations unless they are required by the data. This regularization constraint makes the solution unique and smooth. In addition, unlike the average method above, the solution is legitimate and able to converge multiple-image positions in the source plane when delensing is performed.

In this study, we propose a free-form MAXimum-entropy ReconStruction (MARS) method. The basic maximum-entropy principle is that given a prior, the maximum-entropy distribution best represents the current state of the information, making the fewest assumptions. It has been among the most popular regularization schemes used for ill-posed inversion problems and has been applied to WL and WL+SL mass reconstruction studies (Bridle et al. 1998; Seitz et al. 1998; Jee et al. 2007). The current approach in this paper is similar to the WL+SL study of Jee et al. (2007, hereafter J07) with some important updates. First, our grid represents the mass distribution of the SL system, not the lensing potential. Our experiment shows that the difference makes the optimization of the target function (likelihood+regularization) much more stable than the case where the lensing potential is used. Because the relation between the mass and deflection field is nonlocal, we have to extend our reconstruction field to be slightly larger than the target reconstruction field in order to account for the contribution from the mass outside the target field. Second, we update the prior normalization scheme used for the cross-entropy computation. This reduces the artifact that flattens the mass peak more than we desire. Third, we employ a new optimization engine, which functions reasonably well even in the case where the number of free parameters is huge (~19,600).

This paper is structured as follows. In Section 2, we describe our algorithm for the SL mass reconstruction and error estimation. In Section 3, we introduce the synthetic SL data that we use to evaluate the performance of our algorithm and real observational SL data of A1689. We demonstrate that the algorithm produces high-fidelity mass reconstruction results for the synthetic clusters and also discuss the result for A1689 in Section 4. Section 5 discusses the dependence of the mass reconstruction on the grid resolution and initial condition. We conclude in Section 6. Unless stated otherwise, this paper assumes the cosmology with \( \Omega_m = 0.27, \Omega_\Lambda = 0.73, \) and \( h = 0.72. \)

2. Method

2.1. MARS Algorithm

MARS is based on a free-form source plane minimization with maximum-entropy regularization. Here, we briefly summarize the algorithm unique to our method. For general details of the SL mass reconstruction, readers are referred to excellent review papers (e.g., Kochanek 2006; Kneib & Natarajan 2011).
Gravitational lensing is a nonlinear mapping that transforms the source position $\beta$ to the image position $\theta$ via the following lens equation:

$$\beta = \theta - \alpha(\theta),$$

(1)

where $\alpha$ is the deflection angle. One can compute the deflection angle $\alpha$ via either the derivative of the deflection potential $\Psi$:

$$\alpha = \nabla \Psi$$

(2)

or the convolution of the convergence $\kappa$:

$$\alpha(\theta) = \frac{1}{\pi} \int \kappa(\theta') \frac{\theta - \theta'}{||\theta - \theta'||^2} d^2 \theta',$$

(3)

The convergence $\kappa$ is the dimensionless surface mass density defined as

$$\kappa = \frac{\Sigma}{\Sigma_c},$$

(4)

where $\Sigma_c$ is the (critical) surface mass density. The $\Sigma_c$ is defined as

$$\Sigma_c = \frac{c^2 D_s}{4\pi G D_{ds} D_{dl}},$$

(5)

where $D_{ds}$ is the angular diameter distance between the source and lens, $D_s$ is to the source, and $D_d$ is to the lens.

The first method (obtaining the deflection angle $\alpha$ by taking the derivative of the lensing potential $\Psi$; Equation (2)) has been advocated over the second one (Equation (3)) by some authors (e.g., Seitz et al. 1998; Bradač et al. 2005; Jee et al. 2007), who claim that the second method leads to biased results because the mass distribution outside the reconstruction field can affect the evaluation of the deflection angle (Equation (3)). However, this weakness of the second method can be overcome by simply and sufficiently extending the reconstruction field (as we demonstrate in Section 4.1). In this paper, we choose the second method because our experiment shows that the function minimization is much more stable than when we use the first method, although the field extension increases the number of free parameters by a significant factor.
Both $\kappa$ and $\alpha$ are functions of the source redshift. Because our sources are at different redshifts, we first evaluate them at a reference redshift $z_f$ and then scale the deflection angle for each source according to its redshift as follows:

$$\alpha(z_i) = \alpha(z_f) W(z_f, z_i),$$

where $W(z_f, z_i)$ is the cosmological weight:

$$W(z_f, z_i) = \frac{D(z_f)D(z_i, z_f)}{D(z_i)D(z_f, z_i)}. \tag{7}$$

$D(z)$ is the angular diameter distance to the redshift $z$ while $D(z_A, z_B)$ is between the redshift $z_A$ and $z_B$. In this paper, we set $z_f = 9$.

All multiple images from each source should converge to the same positions in the source plane when “delensed,” and we minimize their scatters:

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(\theta_{i,j,k} - \alpha_{i,j,k}(z) - \beta_k)^2}{\sigma_i^2}, \tag{8}$$

where

$$\beta_k = \frac{1}{J} \sum_{j=1}^{J} (\theta_{i,j,k} - \alpha_{i,j,k}(z)),$$ \tag{9}

and $I$ is the total number of sources, $J$ is the total number of multiple images for each source, and $K$ is the total number of “knots” for each multiple image. Here, “knots” refer to the positions of distinctive morphological features within a source if identifiable, which also must converge if the mass reconstruction is flawless. In this paper, we do not utilize the information (1) because our synthetic SL data do not provide them and (2) because A1689 has only a few sources with clear knots. Nevertheless, as explained below, we set up virtual knots as a convenient way to evaluate local magnifications. The notation $\sigma_i$ in Equation (8) is the maximum scatter of the multiple-image positions from the $i$th source in the source plane and is given by

$$\sigma_i = \sqrt{\sigma_{x,i}^2 + \sigma_{y,i}^2}, \tag{10}$$
where $\sigma_{ij}$ and $\sigma'_{ij}$ are the largest differences in $\beta$ along the $x$- and $y$-axes, respectively. If we had used a fixed value for $\sigma_i$ (for example, $\sigma_i = 0''001$), the mass reconstruction would have been biased toward high-magnification results because a higher-magnification model can make the $\chi^2$ term smaller. Using the current scheme prevents the converging model from favoring a high-magnification solution. Although one can consider a full lensing Jacobian matrix and compute the local magnification with it, in this paper, we implement this by simply creating two virtual knots $0''2$ from the source center in

Table 1 (Continued)

| ID   | B05 8 | R.A. (J2000) | Decl. (J2000) | $z$ | Class   |
|------|-------|--------------|---------------|----|---------|
| 21.1 | 21.1  | 13 11 31.027 | −01 20 45.82 | 1.78 | Silver  |
| 21.2 | 21.2  | 13 11 30.819 | −01 20 44.91 | 1.78 | Silver  |
| 21.3 | 21.3  | 13 11 25.250 | −01 20 11.28 | 1.78 | Silver  |
| 22.1 | 22.1  | 13 11 29.694 | −01 20 08.84 | 1.7 | Gold    |
| 22.2 | 22.2  | 13 11 29.617 | −01 20 23.76 | 1.7 | Gold    |
| 22.3 | 22.3  | 13 11 32.420 | −01 21 15.04 | 1.7 | Gold    |
| 23.1 | 23.1  | 13 11 29.533 | −01 20 10.08 | 2  | Silver  |
| 23.2 | 23.2  | 13 11 29.558 | −01 20 22.06 | 2  | Silver  |
| 23.3 | 23.3  | 13 11 32.662 | −01 21 15.26 | 2.6 | Gold    |
| 24.1 | 24.1  | 13 11 29.192 | −01 20 56.22 | 2.6 | Gold    |
| 24.2 | 24.2  | 13 11 32.059 | −01 19 50.56 | 2.6 | Gold    |
| 24.3 | 24.3  | 13 11 30.290 | −01 19 34.21 | 2.6 | Gold    |
| 24.4 | 24.4  | 13 11 33.719 | −01 20 19.91 | 2.6 | Gold    |
| 26.1 | 26.1  | 13 11 25.153 | −01 20 32.74 | 0.96 | Silver |
| 26.2 | 26.2  | 13 11 31.326 | −01 20 25.28 | 0.96 | Silver |
| 26.3 | 26.3  | 13 11 30.242 | −01 20 32.63 | 0.96 | Silver |
| 27.1 | 27.1  | 13 11 25.173 | −01 20 33.16 | 1  | Silver  |
| 27.2 | 27.2  | 13 11 31.369 | −01 20 24.64 | 1  | Silver  |
| 27.3 | 27.3  | 13 11 30.192 | −01 20 32.94 | 1  | Silver  |
| 28.1 | 28.1  | 13 11 28.301 | −01 20 15.06 | 5.45  | Silver  |
| 28.2 | 28.2  | 13 11 34.262 | −01 20 00.01 | 5.45  | Silver  |
| 29.1 | 29.1  | 13 11 29.226 | −01 20 58.06 | 2.5 | Gold    |
| 29.2 | 29.2  | 13 11 30.022 | −01 19 34.20 | 2.5 | Gold    |
| 29.3 | 29.3  | 13 11 32.177 | −01 19 52.89 | 2.5 | Gold    |
| 29.4 | 29.4  | 13 11 33.626 | −01 20 20.64 | 2.5 | Gold    |
| 30.1 | 30.1  | 13 11 32.420 | −01 19 19.80 | 3.0 | Gold    |
| 30.2 | 30.2  | 13 11 33.185 | −01 19 26.07 | 3.0 | Gold    |
| 30.3 | 30.3  | 13 11 33.662 | −01 19 32.73 | 3.0 | Gold    |
| 31.1 | L07   | 13 11 32.191 | −01 20 03.53 | 3.0 | Gold    |
| 31.2 | L07   | 13 11 33.216 | −01 20 20.90 | 3.0 | Gold    |
| 31.3 | L07   | 13 11 29.589 | −01 21 02.83 | 3.0 | Gold    |
| 31.4 | L07   | 13 11 29.804 | −01 20 43.35 | 3.0 | Gold    |
| 31.5 | L07   | 13 11 28.449 | −01 21 00.66 | 4.58 | Silver  |
| 31.6 | L07   | 13 11 34.653 | −01 20 33.60 | 4.58 | Silver  |
| 31.7 | L07   | 13 11 28.560 | −01 20 59.35 | 1.9 | Gold    |
| 31.8 | L07   | 13 11 33.953 | −01 20 32.52 | 1.9 | Gold    |
| 31.9 | L07   | 13 11 29.431 | −01 20 34.66 | 1.9 | Gold    |
| 35.1 | L07   | 13 11 31.566 | −01 19 45.94 | 3.0 | Gold    |
| 35.2 | L07   | 13 11 31.686 | −01 19 47.39 | 3.0 | Gold    |
| 40.1 | L07   | 13 11 30.260 | −01 20 12.04 | 2.52 | Gold    |
| 40.2 | L07   | 13 11 26.176 | −01 21 03.29 | 2.52 | Gold    |
| 46.1 | C10   | 13 11 31.669 | −01 20 47.19 | 3.48  | Gold Candidate |
| 46.2 | C10   | 13 11 24.959 | −01 20 13.98 | 3.48  | Gold Candidate |
| 50.1 | C10   | 13 11 32.580 | −01 20 43.60 | 4.27  | Gold Candidate |
| 50.2 | C10   | 13 11 31.021 | −01 21 09.08 | 4.27  | Gold Candidate |
| 50.3 | C10   | 13 11 31.660 | −01 20 13.66 | 4.27  | Gold Candidate |

Notes:

8 We follow the numbering scheme of B05.
9 Photometric redshifts.
10 Spectroscopic redshift data from B16.

- The Astrophysical Journal, 931:127 (27pp), 2022 June 1

Cha & Jee

The Astrophysical Journal, 931:127 (27pp), 2022 June 1

Cha & Jee
such a way that the two vectors connecting the knots and the center are perpendicular to each other. Figure 1 illustrates the concept of our virtual knots in the image plane and the way that we evaluate $\sigma_{x,i}$ and $\sigma_{y,i}$ in the source plane.

The following cross-entropy\(^5\) is used to regularize our mass reconstruction to achieve smooth results and prevent overfitting:

$$R = \sum_{m,n} \left[ p(m,n) - \kappa(m,n) + \kappa(m,n) \ln \left( \frac{\kappa(m,n)}{p(m,n)} \right) \right],$$

(11)

where $p(m,n)$ and $\kappa(m,n)$ are the prior and convergence at the position $(m,n)$, respectively. This regularization is different from the one in Jee et al. (2007), where the the first two terms in Equation (11) are dropped and $p$ and $\kappa$ are normalized in such a way that the summation of each becomes unity. We find that this modification reduces the artificial flattening of the mass peak.

Our final target function to minimize is then

$$f = \chi^2 + rR,$$

(12)

where $r$ is a regularization control parameter to adjust the balance between $\chi^2$ and $R$. Choosing too large values leads to oversmoothed results whereas too small values cause overfitting. In the latter case, the resulting mass map is not uniquely determined and shows a number of (unphysical) small-scale features. Because Equation (12), composed of the log-likelihood and prior terms, is based on the Bayesian principle, one may suggest a value, which makes the $\chi^2$ value per component approach unity (e.g., Seitz et al. 1998). However, in our case, because $\sigma_i$ (Equation (10)) in the $\chi^2$ term (Equation (8))) is not a $\sim$68% error estimate but a maximum dispersion, this scheme is not directly applicable. Instead, we examine the relation between $r$ and the resulting mean scatter in the source plane and choose the value in such a way that the mean scatter becomes $\sim0.001$. We discuss the mass map dependence on the $r$ parameter in Section 5.3.

Our mass reconstruction is performed in two steps. In the first one, we embed a 50 $\times$ 50 target reconstruction field in the center of a 70 $\times$ 70 grid, where the 10 pixel margin surrounding the target field is introduced to include the impact of the mass outside the 50 $\times$ 50 target reconstruction field. After the minimization of the function $f$ in the first step, we double the resolution (140 $\times$ 140) and refine the result from the initial 70 $\times$ 70 solution. As explained below, our mass reconstruction is computationally intensive. This two-step implementation is found to speed up the convergence substantially compared to the case where we start with the 140 $\times$ 140 grid from the beginning. We considered different sizes for the margin and found that larger margins do not change the final result significantly. Because the observed multiple images do not lie exactly on the lattice point of the grid, we use bicubic interpolation to compute $\kappa$ at the image positions.

We start the function minimization from a flat prior. If the minimization is finished for the given prior, we restart the minimization after updating the prior with the new convergence map obtained in the previous minimization. This prior needs to be slightly smoothed to prevent overfitting. We applied a Gaussian convolution kernel whose smoothing scale is a $\sigma = 0.6$ (1.2) grid cell for the 70 $\times$ 70 (140 $\times$ 140) grid. This minimization cycle is repeated until there is no improvement.

Because the number of free parameters is extremely large (4900 and 19,600 for the 70 $\times$ 70 and 140 $\times$ 140 grids, respectively) and the function is highly nonlinear, caution is needed when we select a minimization algorithm. We choose the L-BFGS-B algorithm (Byrd et al. 1995; Zhu et al. 1997), which is based on the quasi-Newton method. The algorithm does not require a full Hessian matrix to determine descent directions, and this is a significant merit in our study because the full Hessian matrix evaluation is prohibitively expensive.

2.2. Uncertainties of Mass Maps

It is important to understand the limitation of the SL mass reconstruction in order to interpret the result quantitatively and compare it with other studies. In the LTM approach, it is straightforward to compute the parameter uncertainties because the number of free parameters is small and the system is in general overdetermined. However, because the quoted uncertainties are valid only within the LTM assumptions, one must include systematic errors due to the chosen LTM models when assessing the validity of the result.

In the non-LTM or free-form method, computation of the parameter uncertainties is nontrivial. In particular, if the minimization function does not contain any regularization term, the system is highly underdetermined and thus the solution is not unique. To bypass the difficulty, some authors perform many minimization runs and quote the standard deviation of these solutions. However, because the mean is not one of the valid solutions, it is questionable whether the standard deviation can be used as a reliable uncertainty proxy.

Our MARS free-form method provides a unique solution because the maximum-entropy term makes the system overdetermined. Obviously, this is a significant advantage over other free-from methods and allows us to estimate the parameter (i.e., mass cell) uncertainties from the Hessian matrix. Although the matrix is extremely large (70$^2$ $\times$ 70$^2$ and 140$^2$ $\times$ 140$^2$ for the 70 $\times$ 70 and 140 $\times$ 140 grids, respectively), the computation is needed only once after the convergence is reached. Also, because evaluation of each matrix element is independent, computation of the entire matrix elements can be easily parallelized. This Hessian-based uncertainty estimation is also used in Jee et al. (2007). However, their parameters are potentials and thus complex error propagation using all covariances was needed to obtain mass uncertainties. Our Hessian matrix is a square matrix whose elements are the second derivatives\(^6\) of the function $f$ (Equation (12)). Under the assumption that the posterior of each $\kappa$ value follows the Gaussian distribution, we can use the inverse of the Hessian matrix to obtain the covariance of $\kappa$.

The covariance matrix needs to be rescaled if the $\chi^2$ value (Equation (8)) per multiple image departs from unity when the convergence is reached. We adopted $\sigma = 0.020$ to be our fiducial mean source plane scatter and multiplied the per-image $\chi^2$ value to our covariance matrix. In Section 5.4, we discuss how the uncertainty depends on our choice for the mean source plane scatter.

---

\(^5\) This term is also referred to as the Kullback–Leibler divergence (Kullback & Leibler 1951).

\(^6\) We computed the second derivatives numerically with respect to each grid cell.
We note that the uncertainty map computed in this way has some limitations. First, it is derived under the assumption that the error distribution is Gaussian. Also, the values are valid only for the given prior. Finally, our uncertainty map does not include systematic errors due to the regularization and other method-specific artifacts.

3. Data

We apply MARS to three galaxy clusters. Two of them are the mock clusters from Meneghetti et al. (2017, hereafter M17) and the other is A1689. Testing our algorithm with the mock clusters is important because this allows us to objectively evaluate the performance by comparing the reconstruction with the “truth.” Also, several existing algorithms have participated in the mass reconstruction of the same mock clusters and this enables straightforward comparisons.

Most of the results presented in M17 were submitted before unblinding. Therefore, one may argue that any results published after unblinding should not be evaluated on equal footing. For LTM-based methods, which can utilize the true halo locations, shapes, slopes, etc. as their initial conditions or constraints, the solutions can be optimized or improved after the truth is unblinded. However, when it comes to grid-based free-form methods like MARS, these knobs do not exist and thus in general it is difficult to utilize the unblinded data. For certain free-form methods that produce different results for different runs/realizations, confirmation bias can occur because one can subjectively discard the samples that are distant from the unblinded truth. Because our mass reconstruction produces a quasi-unique solution regardless of the initial condition/assumption, this confirmation bias is irrelevant to our algorithm.

3.1. Mock-Cluster Data

To test our SL reconstruction method, we use the two mock clusters from M17, whose SL data are publicly available. During the review process of the current article, we were informed that the “true” deflection field data released by M17 are not self-consistent with their multiple-image catalogs. Our independent investigation reveals that indeed the scatterers in the source plane measured with their “true” deflection field data are as large as ~0.1% for both clusters. Note that the results in this paper are obtained with the same multiple-image catalogs available to the M17 participants. Here we provide only a brief summary on how these two mock clusters are generated. For details, we refer readers to M17.

The mock galaxy cluster Ares at $z = 0.5$ is created with the semianalytic code MOKA (Giocoli et al. 2012). Ares consists of three components: (1) the main dark matter halo, (2) cluster members, and (3) the brightest cluster galaxy (BCG), which all follow analytic profiles. The SL data are produced under the flat $\Lambda$CDM model with $\Omega_m = 0.272$ and $H_0 = 70.4$ km s$^{-1}$ Mpc$^{-1}$. The projected mass distribution shows many sharp “spikes” at the location of galaxies due to the cuspy centers of the analytic profile halos.

The other mock galaxy cluster Hera at $z = 0.507$ is generated from a high-resolution ($\rho_{DM} = 10^{8} h^{-1} M_{\odot}$) zoom-in resimulation of the initial low-resolution cosmological simulation of Planelles et al. (2014), which assumes a flat $\Lambda$CDM with $\Omega_m = 0.24$, $\Omega_b = 0.04$, $\sigma_8 = 0.8$, and $H_0 = 72$ km s$^{-1}$ Mpc$^{-1}$.

3.2. A1689 Data

The galaxy cluster A1689 at $z = 0.18$ has been the subject of a number of SL studies (e.g., Broadhurst et al. 2005; Zekser et al. 2006; Limousin et al. 2007; Coe et al. 2010; Diego et al. 2015). We compiled multiple images from the following three studies: Broadhurst et al. (2005, hereafter B05), Limousin et al. (2007, hereafter L07), and Coe et al. (2010, hereafter C10). The multiple images of A1689 used in the current study are listed in Table 1, which contains 109 multiple images from 34 source galaxies. We classify them into the “gold,” “silver,” and “gold candidate” samples. The gold images have spectroscopic redshifts. The silver images have only photometric redshifts, which all three papers agree to be bona fide multiple images. The gold candidates are the multiple images whose spectroscopic redshifts were unknown to the authors in the three studies, but later measured by Bina et al. (2016, hereafter B16). There are 73 multiple images from 22 source galaxies in the gold sample, 31 multiple images from 10 source galaxies in the silver sample, and 5 multiple images from 2 source galaxies in the gold candidate sample. We display the multiple-image distribution used in this study in Figure 4, which shows the central 200′ × 200′ region, where we perform the SL mass reconstruction. We follow the numbering scheme of B05 while adopting the coordinates of C10.

4. Result

Here we present our SL mass reconstruction results for Ares, Hera, and A1689. In Section 4.1, we discuss the results for the two mock clusters and compare them with the “truth.” The A1689 result based on the multiple images in the gold sample is described in Section 4.2. We present the results for different choices of multiple images in Appendix B. The plate scales are 5.998 kpc arcmin$^{-1}$ at the redshift ($z = 0.5$) of Ares, 6.043 kpc arcmin$^{-1}$ at the redshift ($z = 0.507$) of Hera, and 2.963 kpc arcmin$^{-1}$ at the redshift ($z = 0.18$) of A1689.

4.1. Mock-cluster Result

4.1.1. Convergence Map Comparison

Figure 5 shows the truth, reconstructed convergence, and uncertainty maps of Ares and Hera. Our reconstruction closely traces the main clumps of the truth. In the case of Ares, the reconstructed convergence map reveals the two main clumps.
located in the NW and SE regions. The small galaxy-scale spikes are missing in our mass reconstruction because their features are smaller than our grid cell. Instead, our convergence shows small “wrinkles,” which we believe are artifacts due to the resolution limit (for example, it is difficult to make two multiple images within a grid cell converge in the source plane).

The reconstructed mass map of Hera also closely follows the truth for the main clumps. We find that the artifacts mentioned for the Ares reconstruction are substantially reduced here because the true mass distribution, on which the SL data are based, is much smoother. In Hera, one noticeable discrepancy between our reconstruction and the truth is the location of the eastern peak, which is shifted south by \( \sim 10 \) kpc with respect to the truth. We attribute the offset to the combined effect of the relatively insufficient number of multiple images around this eastern peak and the regularization.

Our uncertainty map shows that the uncertainty varies from \( \sigma_\kappa \sim 0.1 \) (0.035) to \( \sim 0.45 \) (0.11) for Ares (Hera). The uncertainty distribution is consistent with our expectation that in general the error is large where the convergence is high. Also, the error is somewhat reduced at the location of the multiple images.

M17 present residual \( \frac{\langle \kappa - \kappa_{\text{true}} \rangle}{\kappa_{\text{true}}} \) convergence maps for the nine different SL mass reconstruction algorithms, which were used in their SL modeling comparison project. M17 show that in general the LTM models perform better for both Ares and Hera. This is somewhat expected (1) because the two mock clusters were created under the LTM assumption and (2) because analytic profiles are reasonably good priors in the outer regions where there are no multiple images. In M17, the performance difference between the LTM and free-form methods is greater for Ares, which comprises cuspy parametric halos. For comparison with the results in M17, we also display our residual convergence maps with the same color table and range in Figure 6. The residual map for Ares shows that our mass reconstruction quality is similar to those of the LTM models in the region where the SL data are present. Compared with GRALE, which leads the free-form method in performance in M17, our mass reconstruction gives smaller residual convergence values over a larger area with less large-scale features. In the case of Hera (right panel of Figure 6), our reconstruction is superior to any of the M17 results in terms of the amplitude and uniformity of the residuals (see Figure 10 in M17). All LTM residuals possess galaxy-scale spikes, which

---

**Figure 4.** Multiple-image distribution of A1689. We display the central 200' x 200' region, which is our mass reconstruction field. Orange (white) circles and numbers indicate the positions and IDs of “gold” (“silver”) images. Magenta is used to annotate “gold candidate.” The color-composite image is created with the ACS F475W, F625W, and F775W filters.
we attribute to mismatches between the assumed and true halo shapes. All free-form methods suffer from large-scale residuals. As for the mass estimation of galaxy-scale substructures, MARS cannot constrain them because the current grid-based representation cannot distinguish the projected mass belonging to the cluster halo from those bound to the substructures. In addition, these galaxy-scale substructures are smaller than our grid cell. M17 shows that some LTM-based methods...
outperform free-from algorithms for the substructure mass recovery, which is not surprising when we consider the fact that the LTM methods explicitly include the galaxy-scale halos into their models.

4.1.2. Radial Profile Comparison

We compare the radial convergence and cumulative mass profiles in Figures 7–10. M17 state that the SL data are expected to provide tight constraints at \(20'' < r < 60''\) (\(10'' < r < 30''\)) for Ares (Hera). Thus, we choose the radial ranges for comparison accordingly. Because both Ares and Hera comprise two main halos, we measure the radial profiles twice for each cluster using the two centers of the main clumps. Note that M17 use only the stronger peak (the southeastern peak for Ares and the western peak for Hera) for each cluster to measure the profiles. Overall, both radial convergence and cumulative mass profiles are in excellent agreement with the truth. In the study of M17, the three best-performing LTM methods produce the mass profiles for Ares, which differ from the truth by \(<2\%\), while most free-form methods depart from the truth by \(5\%–15\%\). Our result (see the right panel of Figure 8) shows that the agreement of the cumulative mass profile in the \(20'' < r < 60''\) radial range is at the \(\lesssim 1\%\) level. When we choose the northwestern peak as the reference (note that this measurement is not presented in M17), the difference starts to increase at \(r > 30\) rapidly and reaches \(\sim 10\%\) at \(r = 60''\) (see the left panel of Figure 8). Because the northeastern peak has a smaller (\(\lesssim 30''\)) Einstein radius, this behavior is not surprising. In the case of Hera, the mass profiles agree with the truth within \(\lesssim 1\%\) for the radial range \(10'' < r < 30''\) regardless of the center choice. The best LTM methods in M17 show agreements at the \(\sim 3\%\) level. The free-form results in M17 differ from the truth by \(5\%–10\%\).
Figure 8. Cumulative projected mass profiles of Ares. The rest are the same as the descriptions in Figure 7 except that the horizontal black solid lines indicate ±1% departures.

Figure 9. Radial convergence profiles of Ares. Same as Figure 7 except that the black solid lines in the bottom panels indicate the ±5% differences.

Figure 10. Cumulative projected mass profiles of Hera. The rest are the same as the descriptions in Figure 9 except that the horizontal black solid lines indicate ±1% departures.
4.1.3. Magnification Comparison

In Figure 11, we display the magnification maps for Ares and Hera. Also shown in Figure 11 are the true magnification maps that we computed from the publicly available convergence and shear data. The overall shape and size of the reconstructed critical curves are well recovered, although the largest discrepancies in magnification are found near the critical curves as shown in the residual maps (Figure 12). As noted by M17, the critical curves define the regions where the magnification diverges and thus small differences in the mass reconstruction can cause large magnification errors near them. M17 report that some free-form approach such as the Bradac-Hoag, Coe, and Diego-multires methods produce highly fluctuating irregular critical lines, which are not observed in the truth. As shown, our entropy-based regularization enables us to obtain relatively smooth critical curves. Following M17, we also compute the correlation between our and the truth magnification maps and display the result in Figure 13, which indicates that the accuracy (i.e., departure of the median from the perfect correlation) is within ~10% in the range 1 < μ < 30 whereas the precision (i.e., distance between the 25th and 75th percentiles) is ~20% at μ = 10. Note that some details in our magnification comparison methods may be different from M17. For example, we do not use exactly the same region for comparison. With this caveat, we report that our prediction on the median magnification is accurate within 10% for both Ares and Hera. In the case of Ares, we report that our magnification reconstruction results are significantly better than some (Bradac-Hoag, Diego-multires, Lam, and Coe) of the non-LTM methods while the Diego-reggrid and GRALE methods show a similar performance to ours. Clearly, most LTM methods outperform MARS for the Ares magnification reconstruction. In the case of Hera, the performance of MARS is similar to those of GLAFIC and Diego-overfit, which are the best-performing methods in the LTM and free-form algorithms, respectively. Because the two mock clusters are generated under the LTM assumption, there exists a strong correlation between dark matter and luminous galaxies in the mock data. As noted by M17, it is possible that LTM approaches can benefit from the use of this extra information. M17 show that most LTM methods perform better in the case of Ares than...
Hera, which may hint at this possibility because Ares has tighter dark-matter–galaxy correlations. In addition, the critical curves from the LTM methods are much smoother by construction, which is a significant advantage because even small irregular features around the critical curves in non-LTM methods critically contribute to the degradation of the magnification correlations.

4.2. A1689 Result

Here we focus mainly on the convergence, uncertainty, magnification, and radial profile. In Appendix C and D, we present our reconstruction of the multiple images in the source plane and the giant arcs in the image plane.

4.2.1. Convergence and Uncertainty Maps

We display the reconstructed $\kappa$ distribution and uncertainty map of A1689 in Figure 14. Also, Figure 15 shows the $\kappa$ contour overlaid on the color-composite image of the cluster. Readers are reminded that here we only present the result obtained with the gold sample. We discuss how the result changes when we include the gold candidate or silver samples in Appendix B.

The mass map shows that the projected mass distribution of A1689 is characterized by the dominant mass peak centered on the BCG at $(\alpha, \delta) \sim (13^\text{h}11^\text{m}29^\text{s}, -1^\circ20'30'')$ and the secondary, but much weaker, mass peak centered on the bright galaxy at $(\alpha, \delta) \sim (13^\text{h}11^\text{m}32^\text{s}, -1^\circ20'00'')$. This feature has also been observed in some previous LTM-based studies (e.g.,...
Broadhurst et al. 2005; Halkola et al. 2006; Zekser et al. 2006; Limousin et al. 2007). Some free-form methods do not constrain the secondary peak mentioned above as the characteristic feature of A1689. For example, Coe et al. (2010) obtained a mass map with several substructures whose projected densities are comparable and do not identify our secondary clump as a distinguishing feature of A1689. Diego et al. (2005) presented a result from averaging over 1000 solutions. Although their result shows a weak secondary substructure near our mass peak, it is offset by \( \sim 20'' \), which is surprising if true within the LTM paradigm. We note that in their revised analysis (Diego et al. 2015), this secondary peak coincides with the galaxy.

Given that our mass map is the solution that successfully delenses all multiple images to converge in the source plane with a negligible scatter of \( \sim 0.001 \), its smoothness is remarkable for a free-from result. As already mentioned, the mass reconstructions from Coe et al. (2010) and Diego et al. (2005) suffer from high-amplitude small-scale fluctuations. Mohammed et al. (2014) produce a smooth mass reconstruction of A1689 using the GRALE method, which does not assume LTM. However, because the authors still assume that the cluster consists of Plummer spheres, this smoothness is achieved by design.

We also present the multiple-image scatters in the image plane in Figure 16. Except for the two largest scatters at \( \sim 0.025 \) and \( \sim 0.073 \), we find that most multiple images in the A1689 field have a small \( \langle < 0.01 \rangle \) scatter. We note that these scatters are evaluated without including false multiple images, which would produce catastrophically large scatters. Because MARS is designed to minimize scatters only in the source plane, the model cannot prevent false multiple images. This weakness is shared by the SL reconstruction algorithms that employ only source plane minimization.

Our scrutiny of the mass map indicates that the mass map contains several “pinches” or “quadrupoles” characterized by alternating over- and underdense regions. We believe that they are artifacts caused by the resolution limitation and happen where compact halos generate multiple images densely distributed in small areas. As discussed in Section 4.1, the mass reconstruction for Ares produces many such features while no such artifact is seen in the result for Hera. The uncertainty map (middle panel of Figure 14) somewhat resembles the mass reconstruction (top panel of Figure 14), which is similar to our mock-cluster cases (Section 4.1). Near the central mass peak, this convergence uncertainty \( \langle \delta \kappa \rangle \) is high \( \langle \delta \kappa / \kappa \rangle \sim 0.8 \) whereas near the field boundary it is low \( (0.3 \lesssim \delta \kappa \lesssim 0.4) \). When rescaled by the local convergence, the relative \( \langle \delta \kappa / \kappa \rangle \) error map (bottom panel) shows the opposite trend, with the estimate being high \( \langle \delta \kappa / \kappa \rangle \sim 1 \) near the field edges, which is consistent with the relative difference map shown for the mock cluster (Figure 6). Note that the uncertainties are relatively low at the location of the multiple images.

4.2.2. A1689 Magnification

Figure 17 presents our critical curves of A1689. The global morphology comprising the inner (radial) \( r \sim 15'' \) and outer (tangential) \( r \sim 40'' \) critical curves is similar to the findings in previous studies. However, details differ greatly among the
studies, including ours. Therefore, despite the fact that A1689 is one of the most studied SL clusters, the community still needs to converge on the magnification details and thus the interpretation of high-\(z\) galaxies found behind the cluster. We repeat that because the relation between mass and magnification distributions are highly nonlinear, even small differences in the convergence lead to large changes in the magnification.

As seen in the mock-cluster results, the smoothness of our critical curve is noteworthy when compared with other free-form results. For example, because no regularization is imposed in Coe et al. (2010), their critical curves show highly oscillatory features and vary widely among different solutions, which, however, are all capable of converging all used multiple-image positions in the source plane. In Diego et al. (2005), only the critical curve estimated from the mean (out of 1000) convergence map is shown.

4.2.3. A1689 Radial Profile

The experiments with the mock-cluster SL data (Section 3.1) show that our mass reconstruction recovers the cluster radial profiles with superb (<1\%) accuracy in the regions where multiple images are present. Thus, if all multiple images are real in the gold sample, we expect that in the current study a similarly high-fidelity mass profile can be obtained for A1689 in the SL regime.

Our test with the mock-cluster data shows that our mass reconstruction recovers well the central profile down to \(r \approx 4''\) (\(\sim 25\) kpc). At smaller radii, the density becomes nonnegligibly underestimated because of the limitations in resolution and regularization. Considering the plate scale and resolution of our mass reconstruction for A1689, we believe that the smallest radius at which we can restore the density without severe bias is \(\sim 12\) kpc. Given the extent of the multiple images, the mass density is expected to be constrained by the data out to \(\sim 240\) kpc.

We display the projected radial density profile of A1689 in the top panel of Figure 18. Also plotted is the best-fit (\(c_{200} = 5.53 \pm 0.77\), and \(r_s = 538^{+90}_{-100}\) kpc) NFW profile. This NFW profile is a reasonable approximation of the overall shape of our radial profile in the SL regime, although we observe some indications of the densities predicted by the best-fit NFW model being 1\(\sigma\) higher at \(r \lesssim 15\) kpc and lower.

Figure 15. Mass contours overlaid on the ACS color-composite image. The white contours and labels show the convergences. The convex hull (solid pink) encloses the area where multiple images are located. The color-composite image is created with the F475W (blue), F625W (green), and F775W (red) filters. The projected mass distribution of A1689 is characterized by the primary mass peak centered on the BCG at \((\alpha, \delta) \approx (13^\h 1^m 29^s, -1^\d 20^\prime 30^\prime)\) and the much weaker secondary mass peak centered on the bright galaxy at \((\alpha, \delta) \approx (13^\h 1^m 32^s, -1^\d 20^\prime 00^\prime)\).
at \( r \gtrsim 200 \) kpc. Because there exist several multiple images at \( r \lesssim 15 \) kpc, it is unlikely that the shallower (than the best-fit NFW prediction) profile is entirely due to the aforementioned artifact, but is somewhat constrained by the data. The “shoulder” (see the arrow in the top panel) at \( r \sim 160 \) kpc coincides with the location of the secondary mass peak at \((\alpha, \delta) = (13^\circ 11^\prime 02^\prime, -1^\circ 20^\prime 00^\prime)\), which indicates that our best-fit NFW model might be nonnegligibly affected by the presence of the substructure. The cumulative projected mass profile (the bottom panel of Figure 18) shows that the aperture mass of A1689 is \( \sim 4 \times 10^{14} M_\odot \) at \( r = 237 \) kpc, which is in good agreement with the previous studies (e.g., Broadhurst et al. 2005; Halkola et al. 2006; Limousin et al. 2007). At face value, the best-fit concentration value \( c_{200} = 5.53 \pm 0.77 \) is slightly lower than the results in some previous studies. However, when we limit our comparison to those who used only the SL data (e.g., Broadhurst et al. 2005; Limousin et al. 2007; Coe et al. 2010), our measurement is consistent at the \( \sim 1\sigma \) level.

5. Discussion

5.1. Effect of Grid Resolution

As mentioned in Section 2.1, our final reconstruction is achieved in two steps (i.e., initial reconstruction in the \( 50 \times 50 \) grid followed by a second reconstruction in the \( 100 \times 100 \) grid\(^8\)).

\(^8\) These refer to the resolutions when the margins are excluded.
Figure 19. Resolution test. The left panels (right panels) show the mass reconstructions with the $50 \times 50$ ($100 \times 100$) grid. The $100 \times 100$ versions reduce the quadrupole-like artifacts mentioned in Section 4.2.1. In particular, artifact reduction is most significant in Ares. Apart from the artifact reduction, no significant systematic difference is found between the two resolution cases.
Figure 20. Dependence of mass reconstruction on initial conditions. We experiment with four different initial $\kappa$ distributions. Upper left: a flat mass sheet of $\kappa = 0.5$. Upper right: a flat mass sheet of $\kappa = 0.1$. Middle left: an NFW profile from Coe et al. (2010) ($c_{200} = 7.6$, $r_s = 338$ kpc). Middle right: an NFW profile from this work ($c_{200} = 5.53$, $r_s = 538$ kpc). To guide eyes, we use black dots to mark the locations of the multiple images. The mass map comparison shows that the agreement among the four results is excellent in the region constrained by the multiple images. Bottom: comparison of the four radial profiles. The mass map comparison shows that the agreement among the four results is excellent in the region constrained by multiple images. No significant difference is found out to $r \sim 200$ kpc. Beyond this radius, some small systematic differences are noticeable, which is not surprising because the SL data do not provide any meaningful constraint there. The reference redshift is $z_f = 9$. 

The Astrophysical Journal, 931:127 (27pp), 2022 June 1

Cha & Jee
If significant systematic differences are found between the two results, this might indicate that the resolution in the second step is insufficient and the result should be further improved with even a higher-resolution grid. Here we demonstrate that although the larger grid provides a higher resolution in some cases, no significant systematic difference is found.

Figure 19 compares the 50 × 50 and 100 × 100 versions side by side for the three clusters studies in this paper. It is clear that the larger grid improves the resolution of the resulting mass map. In addition, the 100 × 100 versions reduce quadrupole-like artifacts mentioned in Section 4.2.1. In particular, the artifact reduction is most significant in Ares, whose truth map possesses many sharp spikes, which cannot be sufficiently represented even with the 100 × 100 resolution.

We also notice that the convergence values at the mass peaks are slightly reduced in the higher-resolution version. We speculate that this might arise from our two-step mass reconstruction method, which obtains the lower-resolution version first and uses the smoothed version as a prior to the second reconstruction. However, the differences make only a negligible impact on the radial mass profiles.

Therefore, we conclude that although the higher-resolution grid improves the resolution and reduces some artifacts, the change in grid resolution does not lead to any significant systematic differences that lead to changes in scientific interpretation.

5.2. Effect of Initial Conditions

In order to show that our algorithm provides quasi-unique solutions, we demonstrate that the mass reconstruction does not depend on our initial guesses. In Figure 20, we present the A1689 mass reconstruction results when we start the function optimization from four different initial conditions: two flat mass sheets and two NFW profiles.

The 2D mass map comparison shows that the agreement among the four results is excellent in the region constrained by multiple images (to guide eyes, we plot the locations of the multiple images with black dots). The mass map in the outskirts where no SL data exist somewhat depends on the initial conditions, which however is not surprising because no constraints are provided there.

For a more quantitative analysis, we compare the radial profiles from these four mass maps in the bottom panel of Figure 20. As indicated by the 2D comparison, no significant difference is found out to r ∼ 200 kpc. Beyond this radius, some small systematic differences are noticeable. In Figure 21, we display the residual mass maps.

In summary, we find that MARS achieves a quasi-unique solution for the region constrained by the SL data. We note that the optimization requires a much longer time if the initial setups are far from the final result (e.g., flat initial conditions).

5.3. Effect of Regularization Control Parameter

Equation (12) consists of the χ² and regularization terms. The regularization term is to prevent overfitting and stabilize the convergence to a quasi-unique solution. In order to illustrate the dependence of the mass map on the regularization control parameter r, we display the three (r = 0, 0.1, 1, and 10) cases in Figure 22. The r = 1 case corresponds to our fiducial result, which leads to a mean source plane scatter of ∼10⁻³μ. The r = 0.1 (10) results yield a mean source plane scatter of ∼10⁻⁴μ (∼10⁻²μ). The differences among the r = 0.1, 1, and 10 cases are small, but noticeable. As expected, the smaller (larger) r value produces a less (more) smoothed mass map.

The r = 0 mass map is obtained without any entropy prior and it is not surprising that the resulting mass map shows a number of unphysical, small-scale fluctuations. Because the system is underdetermined, the solution is not unique and the result displayed in Figure 22 is just one of many possible solutions.

5.4. Dependence of Error Estimation on Source Position Scatters

As explained in Section 2.2, we estimate the mass reconstruction errors with the Hessian matrix by assuming that the posteriors follow Gaussian distributions. Another important assumption is the size of the multiple-image position scatters in the source plane. For our fiducial result, we assume σ = 0°0.02 after considering the finite resolution of our mass grid. Because it is difficult to define/estimate the exact value for this scatter from the first principle, here we examine the dependence of the mass reconstruction error estimation on the size of the scatter.

In Figure 23, we show the uncertainty maps of Hera and Ares for the three additional cases: σ = 0°0.05, 0°0.05, and 0°5. We find that the relative scales across the field are not affected by the assumed scatter values, which is somewhat expected because changing the scatter is equivalent to scaling the likelihood with a
simple multiplicative factor. In terms of the absolute values, we notice significant changes. For Ares, the decrease in error is nearly an order of magnitude when we change $\sigma = 0''005$ to $\sigma = 0''05$.

For the case of $\sigma = 0''5$, the decrease is about 50% relative to the $\sigma = 0''05$ case. The sensitivity is reduced for Hera, and we attribute this to the difference in the number of multiple images.

Figure 22. Dependence of the reconstruction on the regularization control parameter $r$. We display the A1689 results obtained with different $r$ values. Upper left: result with $r = 1$ (fiducial). Upper right: result with $r = 0$ (no regularization). Middle left: result with $r = 0.1$. Middle right: result with $r = 10$. In the bottom panel, we display the residual maps for the $r = 0.1$ and 10 cases. The reference redshift is $z_f = 9$. 
In Figure 24, we show the relative uncertainty ($\delta\kappa/\kappa$) maps of Hera and Ares for the three cases. Similarly to Figure 23, the uncertainty values scale with the size of the source plane scatter.

6. Conclusion

We have presented a new free-form maximum-entropy-regularized SL mass reconstruction method (MARS). The target function that we minimize consists of two parts: log-likelihood and regularization terms. The log-likelihood term is a chi-square function of multiple-image positions in the source plane whereas the regularization term is a cross-entropy between the current mass map and the prior. The mass model is represented by a square grid of convergence values, and we perform the mass reconstruction in two steps: initial reconstruction with a $50 \times 50$ ($70 \times 70$ with the margin) resolution followed by a
100 × 100 (140 × 140 with the margin) resolution. The method enables us to converge all legitimate multiple-image positions in the source plane, suppress spurious small-scale fluctuations due to overfitting, and provide a quasi-unique solution independently of initial conditions.

We validate our algorithm with the publicly available synthetic SL data FF-SIMS, which have been used to test a number of existing algorithms. We find that the overall performance is on par with the results from the best-performing LTM methods published in the literature, which is remarkable because the synthetic SL data were generated under the LTM assumption, and most non-LTM methods have been shown to perform relatively poorly by a significant margin. In particular, our mass profile excels in the mass profile reconstruction,

Figure 24. Dependence of the relative uncertainty estimation on the source scatter. The left (right) panels show the relative uncertainty maps of Ares (Hera). Top: \( \sigma = 0.005" \). Middle: \( \sigma = 0.05" \). Bottom: \( \sigma = 0.5" \).
agreement with the truth at the subpercent level in the region constrained by the SL data; the best-performing LTM methods is known to achieve ∼2% level accuracy whereas most non-LTM methods deviate from the truth by 5%–15%.

For the reconstruction of galaxy-scale substructures, our algorithm underperforms compared to the LTM methods. Also, in the case of Ares, the magnification prediction from the the LTM methods is found to be superior to our method. These performance differences might be attributed to the possibility that the LTM approaches can benefit from the use of LTM hypothesis, which was also employed in the synthetic cluster generation.

We apply our method to A1689, which is one of the most studied SL clusters. We find that the cluster mass in the SL regime is dominated by the primary halo centered on the BCG and the weaker secondary halo at the bright cluster member ∼160 kpc northeast of the BCG. The A1689 radial profile is well described by an NFW profile with c_{200} = 5.53 ± 0.77 and r_{s} = 538_{-100}^{+90} kpc, which provides no evidence for over-concentration when only the SL regime is considered.

This work is based on observations created with NASA/ESA Hubble Space Telescope and downloaded from the Mikulski Archive for Space Telescope (MAST) at the Space Telescope Science Institute (STScI). The current research is supported by the National Research Foundation of Korea under program 2022R1A2C1003130.

Software: Astropy (Astropy Collaboration et al. 2013), Matplotlib (Hunter 2007), NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020).

Appendix A
Multiple-image Catalog of A1689

Table 1 shows 109 multiple images from 34 source galaxies considered in this study. They are classified into the “gold,” “silver,” and “gold candidate” samples. The objects in the gold sample have spectroscopic redshifts. The silver images have only photometric redshifts and are considered bona fide multiple images by B05, L07, and C10. The spectroscopic redshifts of the gold candidate galaxies are unknown to B05, L07, and C10, but are later measured by B16.

Appendix B
Mass Reconstruction of A1689 with Additional Multiple Images

The fiducial result for A1689 is obtained with the “gold” samples (Section 3.2). Here we investigate how the result changes when we consider “silver” and “gold candidate” samples. With the selection of the “gold” + “gold candidate” samples, we are able to successfully delens all multiple images to converge in the source plane. The resulting mass map (Figure 25(a)) is also very similar to our fiducial case. When we use all three samples (“gold” + “gold candidate” + “silver”), we find that the two sources (IDs 26 and 27) in the “silver” sample cannot be delensed to converge in the source plane. We suggest three possibilities, which are not mutually exclusive. First, the two multiple systems are very close to each other on the image plane (see Figure 4). Thus, the resolution of our grid might not be sufficiently large to distinguish the two systems. Second, we might have used inaccurate redshifts for these images. As we mentioned in Section 3.2, the “silver” multiple images have only photometric redshifts. Thus, it is a legitimate concern that for some sources their photometric redshift might be catastrophically different from the true values (Skelton et al. 2014). Third, these multiple images may not be real multiple images. Diego et al. (2015) state that the authors removed sources 26 and 27 from their catalog because the images show different colors in their newly added IR data. Nevertheless, we find that the resulting mass reconstruction (Figure 25(b)) is still similar to our fiducial version, although some additional small-scale “wrinkles” are introduced. Finally, we carry out mass reconstruction using the “gold” + “gold candidate” + “silver” without these two problematic sources (Figure 25(c)) and find that all multiple images are successfully delensed within ∼0.001 in the source plane. Again, as expected, the resulting mass map is in good agreement with the previous cases. In Figure 26, we show the residual maps with respect to the fiducial result. The positions of multiple images also are marked. We can clearly see that the region where multiple images are located shows larger differences from the fiducial one.
Appendix C
Source Image Reconstruction with the A1689 Mass Model

One of the useful applications of the SL mass reconstruction results is the source plane image reconstruction, which not only provides the model verification but also enables us to use SL clusters as a cosmic lens for observations of the more distant universe. In this appendix, we present source image reconstruction of the multiple images in A1689.

Figure 27 displays the reconstructed images of the four sources: IDs 10, 11, 16, and 24. Although in principle, we can apply this delensing to any multiple-image system, these four

Figure 25. Mass reconstruction of A1689 with different source selections. Fiducial: reconstruction with the “gold” sample. (a) Reconstruction with the “gold” + “gold candidate” samples. (b) Reconstruction with the “gold” + “gold candidate” + “silver” samples. (c) Same as (b) except source IDs 26 and 27 are removed. The reference redshift is $z_f = 9$.

Figure 26. Residual mass maps between the three different reconstructions in Figure 25 and the fiducial one. Left: a residual for the mass map with the “gold” + “gold candidate” samples. Middle: a residual for the mass map with the “gold” + “gold candidate” + “silver” samples. Right: a residual for the mass map with same as middle one except source IDs 26 and 27 are removed. Markers indicate the position of additional multiple images. Black triangles (dots) show the gold candidate (silver) multiple images.
Figure 27. Source plane image reconstruction. We display the results for the four sources: IDs 10, 11, 16, and 24, which have a distinctive morphology in the deep HST/ACS F814W image. For each ID, the image on the left-hand (right-hand) side is the observed (reconstructed) image. North is up, and east is left.
sources have a distinctive morphology in the deep F814W image. Among our mass models, we choose the deflection angle obtained from the “gold” + “gold candidate” + “silver” samples (i.e., the model shown in Figure 25(b)). We find that overall the reconstructed images from the same sources show reasonable agreements in morphological features, orientations, parities, and sizes, which is remarkable because our reconstruction is only based on the source center position convergence.

Appendix D
Giant-arc Reproduction with the A1689 Mass Model

In principle, a perfect lens model should be able to reproduce all observed SL features. However, in general, it is difficult to predict the morphology of the giant arcs precisely because their positions are close to the critical curves and thus the result is highly sensitive to small perturbations. Because MARS is a grid-based free-form algorithm, it is not best poised to perform arc reproduction because of its limited resolution. Nevertheless, it would be interesting to investigate how well MARS can reproduce the giant arcs in the A1689 field.

In Figure 28, we display our reconstruction of the two giant arcs, which have a distinct morphology in the deep F814W image. The first arc is from system 13 whereas the second arc consists of systems 8 and 19. The overall extent and morphology of the two arcs resemble the observed ones. Given that MARS cannot include galaxy-scale substructures and the two arcs are positioned near the bright galaxies, we believe that the resemblance, although it is not perfect, is rather remarkable. We stress that MARS only minimizes the source plane scatter without using the image plane morphology of the multiple images.

Figure 28. Giant-arc reproduction in the A1689 field. We choose the two giant arcs that have a distinct morphology in the deep HST/ACS F814W image. The images on the left-hand (right-hand) side are the reproduced (observed) arcs. Although not perfect, the resemblance is remarkable, especially given that MARS only uses the source plane scatter as constraints without utilizing the image plane morphology.
ORCID iDs
Sangjun Cha  https://orcid.org/0000-0001-7148-6915
M. James Jee  https://orcid.org/0000-0002-5751-3697

References
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bartelmann, M., & Schneider, P. 2001, PhR, 340, 291
Bina, D., Pelló, R., Richard, J., et al. 2016, A&A, 590, A14
Bradač, M., Schneider, P., Lombardi, M., & Erben, T. 2005, A&A, 437, 39
Brüggen, M., Hobson, M. P., Lasenby, A. N., & Saunders, R. 1998, MNRAS, 299, 895
Broadhurst, T., Benítez, N., Cole, D., et al. 2005, ApJ, 621, 53
Byrd, R., Lu, P., Nocedal, J., & Zhu, C. 1995, SIAM J. Sci. Comput., 16, 1190
Coe, D., Benítez, N., Broadhurst, T., & Moustakas, L. A. 2010, ApJ, 723, 1678
Deigo, J. M., Broadhurst, T., Benitez, N., et al. 2015, MNRAS, 446, 683
Diego, J. M., Sandvik, H. B., Protopapas, P., et al. 2005, MNRAS, 362, 1247
Finner, K., Jee, M. J., Golovich, N., et al. 2017, ApJ, 851, 46
Giocoli, C., Meneghetti, M., Bartelmann, M., Moscardini, L., & Boldrin, M. 2012, MNRAS, 421, 3343
Halkola, A., Seitz, S., & Pannella, M. 2006, MNRAS, 372, 1425
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hoekstra, H., Franx, M., & Kuijken, K. 2000, ApJ, 532, 88
Hunter, J. D. 2007, CSE, 9, 90
Jee, M. J., Ford, H. C., Illingworth, G. D., et al. 2007, ApJ, 661, 728
Jee, M. J., White, R. L., Ford, H. C., et al. 2006, ApJ, 642, 720
Jullo, E., Kneib, J. P., Limousin, M., et al. 2007, NJPh, 9, 447
Kaiser, N., & Squires, G. 1993, ApJ, 404, 441
Kassiola, A., & Kovner, I. 1993, ApJ, 417, 450
Kim, J., Jee, M. J., Hughes, J. P., et al. 2021, ApJ, 923, 101
Kneib, J.-P., & Natarajan, P. 2011, A&ARv, 19, 47
Kochanek, C. S. 2006, Gravitational Lensing: Strong, Weak and Micro (Berlin: Springer), 91
Kollberg, S., & Leibler, R. A. 1951, Ann. Math. Statist., 22, 79
Limousin, M., Richard, J., Jullo, E., et al. 2007, ApJ, 668, 643
Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97
Medezinski, E., Broadhurst, T., Umetsu, K., et al. 2007, ApJ, 663, 717
Meneghetti, M., Natarajan, P., Coe, D., et al. 2017, MNRAS, 472, 3177
Mohammed, I., Liesenborgs, J., Saha, P., & Williams, L. L. R. 2014, MNRAS, 439, 2651
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Planelles, S., Borgani, S., Fabjan, D., et al. 2014, MNRAS, 438, 195
Schrabback, T., Bocquet, S., Sommer, M., et al. 2021, MNRAS, 505, 3923
Seitz, S., Schneider, P., & Bartelmann, M. 1998, A&A, 337, 325
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Umetsu, K., Broadhurst, T., Zitrin, A., Medezinski, E., & Hsu, L.-Y. 2011, ApJ, 729, 127
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Zekser, K. C., White, R. L., Broadhurst, T. J., et al. 2006, ApJ, 640, 639
Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. 1997, ACM Trans. Math. Softw., 23, 550
Zitrin, A., Broadhurst, T., Umetsu, K., et al. 2009, MNRAS, 396, 1985