Mon. Not. R. Astron. Soc. **362,** 1247–1258 (2005) doi:10.1111/j.1365-2966.2005.09372.x

Non-parametric mass reconstruction of A1689 from strong lensing data with the Strong Lensing Analysis Package

J. M. Diego,[1,2]★ H. B. Sandvik,[3] P. Protopapas,[4] M. Tegmark,[1,2] N. Benítez[5,6] and T. Broadhurst[7]

[1]*MIT Center for Space Research, Cambridge, MA 02138, USA*
[2]*University of Pennsylvania, 209S, 33rd St. Department of Physics and Astronomy, Philadelphia, PA 19104, USA*
[3]*Max-Planck-Institut für Astrophysik, D-85748 Garching, Germany*
[4]*Harvard-Smithsonian Center for Astrophysics, 60 Graden St. Cambridge, MA 02138, USA*
[5]*Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA*
[6]*Instituto de Astrofísica de Andalucía (CSIC). Camino Bajo de Huítor, 24, Granada, Spain*
[7]*School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel*

Accepted 2005 June 29. Received 2005 June 27; in original form 2004 December 28

ABSTRACT

We present the mass distribution in the central area of the cluster A1689 by fitting over 100 multiply lensed images with the non-parametric Strong Lensing Analysis Package. The surface mass distribution is obtained in a robust way, finding a total mass of $0.25 \times 10^{15}\ h^{-1}\ \mathrm{M}_{\odot}$ within a 70-arcsec circle radius from the central peak. Our reconstructed density profile fits well a Navarro–Frenk–White (NFW) profile with small perturbations due to substructure, and is compatible with the more model-dependent analysis of Broadhurst et al. based on the same data. Our estimated mass does not rely on any prior information about the distribution of dark matter in the cluster. The peak of the mass distribution falls very close to the central dominant (cD) galaxy and there is substructure near the centre suggesting that the cluster is not fully relaxed. We also examine the effect on the recovered mass when we include the uncertainties in the redshift of the sources and in the original shape of the sources. Using simulations designed to mimic the data, we identify some biases in our reconstructed mass distribution. We find that the recovered mass is biased toward lower masses beyond 1 arcmin (150 kpc) from the cD galaxy and that in the very centre we may be affected by degeneracy problems. On the other hand, we confirm that the reconstructed mass between 25 and 70 arcsec is a robust, unbiased estimate of the true mass distribution and is compatible with an NFW profile.

Key words: methods: numerical – galaxies: clusters: general.

1 INTRODUCTION

The breathtaking image of A1689 captured by the Advanced Camera for Surveys (ACS) on-board the *Hubble Space Telescope* (*HST*; Broadhurst et al. 2005a, hereafter B2005) provides us with an unprecedented number of strong lensing arcs in a single cluster. The large number of arcs is due to a combination of deep multicolour imaging with the *HST* and the inherently large Einstein radius of A1689. A total of 106 multiply lensed images of 30 background galaxies have been identified (B2005) and are spread fairly uniformly over an area of diameter ∼300 kpc. In principle, we may obtain an estimate of the deflection angle of the light at the location of each of the images belonging to multiply imaged sources. This deflection relates to the projected gradient of the gravitational potential of the lens, and hence we may derive the surface mass density with a precision and resolution set by the number of multiply imaged sources.

Previous analyses of strong lensing have involved only an order of magnitude fewer arcs per cluster, and hence have not permitted the application of a non-parametric approach, leading to only model-dependent statements in general (Kneib et al. 1993, 1995, 1996; Tyson, Kochanski & Dell'Antonio 1998; Broadhurst et al. 2000; Sand, Treu & Ellis 2002; Gavazzi et al. 2004). These models have produced reliable results for simple symmetric situations where the mass enclosed within the Einstein radius is fairly robust to other parameters. The quality of deep images taken with the ACS opens the way to estimating the surface mass distribution directly without resorting to parametric models. Non-parametric approaches have been previously explored in several papers

★E-mail: jdiego@space.mit.edu

(Saha & Williams 1997; Abdelsalam, Saha & Williams 1998a,b; Trotter, winn & Hewitt 2000; Williams & Saha 2001; Warren & Dye 2003; Bradac et al. 2005; Saha & Williams 2004; Treu & Koopmans 2004) and more recently in Diego et al. (2005, hereafter Paper I). In Paper I, we showed that it is possible to non-parametrically reconstruct a generic mass profile (with substructure) provided the number of arcs with known redshifts is sufficiently large. We have also shown how working with extended images, rather than just their positions, adds enough constraints to the problem, thus solving the regularization problem found in other non-parametric algorithms (see Kochanek, Schneider & Wambsganss 2004 for a discussion on this issue).

The mass distribution of A1689 has recently been estimated using a flexible parametric approach by B2005, who have identified over 100 background galaxies using their method. This analysis assumed a smooth dark matter component for the bulk of the mass in the cluster plus a small lumpy component of mass corresponding to the cluster sequence galaxies. The cluster galaxy contribution is allowed freedom in the ratio of M/L, but the smooth component is fitted to a low-order two-dimensional polynomial, the coefficients of which were optimized to fit the multiply imaged sources. The model is refined as more multiply imaged sources are identified by the model and incorporated to improve its accuracy. Using this approach, B2005 have been able to reliably uncover 106 multiply lensed images of 30 background galaxies.

Non-parametric methods are interesting to explore as they provide a model-independent estimate of the mass distribution, free of assumptions regarding the distribution of mass in the lens plane. Hence, this method provides a very important consistency check, which should be carried out in addition to, but not necessarily at the expense of, parametric methods. If the recovered mass distribution concurs with parametric estimates, this will add to the credibility of these results. If, on the other hand, there are significant deviations, this should open the door to interesting debates trying to understand them. Another major advantage of the non-parametric method is that it allows us to estimate the systematics and errors in the recovered mass distribution free of model assumptions. As shown in Paper I, the minimization process can take as little as a few seconds, which allows for multiple minimizations with random initial conditions. We can then study the dispersion of the recovered solution and consequently provide an error estimate.

In this paper we use one of the algorithms in the Strong Lensing Analysis Package (SLAP) developed by us and introduced in Paper I to reconstruct the mass distribution of A1689. We therefore start by giving a brief summary of the main ingredients of the method.

2 METHOD

2.1 Mass–source inversion of the data

The method described in this section is based on Paper I, and the interested reader is highly encouraged to consult that paper for the finer details. This section simply highlights the main ingredients although with an important difference, which will be discussed at the end of this section.

The problem we want to solve is the inversion of the lens equation

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}, M) \tag{1}$$

where $\boldsymbol{\beta}$ are the unknown positions (β_x, β_y) of the background galaxies, $\boldsymbol{\theta}$ are the observed positions (θ_x, θ_y) of the lensed galaxies (arcs) and $\boldsymbol{\alpha}(\boldsymbol{\theta})$ is the deflection angle created by the lens, which depends on the observed positions, $\boldsymbol{\theta}$, and the unknown mass distribution of the cluster, M. The unknowns of the problem are then $\boldsymbol{\beta}$ and M.

Due to the (non-linear) dependency of the deflection angle, $\boldsymbol{\alpha}$, on the position in the sky, $\boldsymbol{\theta}$, the problem is usually regarded as a typical example of a non-linear problem. However, the problem also has an equivalent formulation which can be expressed in a linear form. The linearization of the problem is possible due to an observational constraint and a fundamental principle.

The constraint is that the observation fixes the positions of the arcs, $\boldsymbol{\theta}$. The non-linear nature of the problem is associated only with this variable. Fixing $\boldsymbol{\theta}$, transforms this variable into a constant.

The fundamental principle is the linear nature of the gravitational potential. The integrated effect of the continuous mass distribution in α can be approximated by a superposition of discretized masses. The continuous mass distribution can be discretized into small cells in the lens plane if the continuous mass distribution can be approximated by a constant over each one of the individual cells or, in other words, if the continuous mass distribution does not change much over the scale of the cells. This can be achieved if we divide the lens plane into a multiresolution grid where the size of the cell in a given position is inversely proportional to the mass density in that position. Other interesting possibilities can be explored. For instance, one may prefer to sample better the regions in the lens plane which are near the arcs where the effects of the finer details in the mass distribution are more evident. This would provide a grid which is constant from iteration to iteration. One may also think about a combination of the two (i.e. sampling proportional to the quantity $M_{\rm cell}/\theta$). This would still be an adaptive process, but takes into account that it is in fact the mass multiplied by $1/\theta$, which determines the lensing effect. These possibilities have not yet been explored.

Using a multiresolution grid with $N_{\rm c}$ cells, and with positions of the arcs, $\boldsymbol{\theta}$, fixed by observations, the problem can be rewritten in the linear algebraic form

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \Upsilon \boldsymbol{M}, \tag{2}$$

where $\boldsymbol{\theta}$ is a vector with $2N_\theta$ elements containing all the observed positions (x and y) of the N_θ pixels in the arcs of the lensed galaxy (or galaxies if there is more than one source), $\boldsymbol{\beta}$ is made up of the corresponding $2N_\theta$ positions (x and y) of the source galaxy, and Υ is a matrix of dimension $2N_\theta \times N_{\rm c}$ where $N_{\rm c}$ is the number of cells of the multiresolution grid used to divide the lens plane.

The reader should note that, strictly speaking, our non-parametric algorithm formulated in this way may also be seen as a parametric one, with the parameters being the number of cells of different sizes, $N_{\rm c}$, the location of the different cells in the lens plane, their masses and the position of the sources. The role of the adaptive grid in this context is to fix the location of the cells in the lens plane based on a previous estimate of the solution.

To invert the strong lensing data, we use the algorithm of SLAP which is based on the bi-conjugate gradient method (Press et al. 1997). Instead of solving equation (2) we solve the following

$$\boldsymbol{\theta} = \Gamma \boldsymbol{X}. \tag{3}$$

Here Γ is a matrix of dimension $2N_\theta \times (N_{\rm c} + 2N_{\rm s})$ and $\boldsymbol{X}$ is the vector of dimension $(N_{\rm c} + 2N_{\rm s})$ containing all the unknowns in our problem, the $N_{\rm c}$ cell masses, M, and the $2N_{\rm s}$ central positions, $\beta_{\rm o}$ (x and y), of the $N_{\rm s}$ sources. From now on we will drop the vector notation.

The bi-conjugate gradient algorithm solves a system of linear equations

$$Ax = b \tag{4}$$

by minimizing the function

$$f(x) = c - bx + \frac{1}{2}x^t Ax, \tag{5}$$

where c is an arbitrary constant. When the function $f(x)$ is minimized, its gradient ($\nabla f(x) = Ax - b$) is zero. The problem is formulated like this because, in most cases, finding the minimum of equation (5) is much easier than finding the solution of the system in equation (4), especially when no exact solution exists for equation (4) or A does not have an inverse.

The algorithm assumes that the matrix A is square. This does not generally hold for our case as, for the matrix Γ, we typically have $N_\theta \gg (N_c + N_s)$. We therefore build a new quantity called the square of the residual, R^2

$$\begin{aligned} R^2 &= (\theta - \Gamma X)^{\mathrm{T}}(\theta - \Gamma X) \\ &= 2\left(\frac{1}{2}\theta^{\mathrm{T}}\theta - \Gamma^{\mathrm{T}}\theta X + \frac{1}{2}X^{\mathrm{T}}\Gamma^{\mathrm{T}}\Gamma X\right). \end{aligned} \tag{6}$$

This is clearly of the same form as equation (5), with $\Gamma^{\mathrm{T}}\Gamma$ a square matrix. Solving the lens equation means minimizing this quantity. The quantity R^2 reaches its minimum in the solution of equation (3), which is also solution of equation (4). We only have to realize that

$$b - AX = \Gamma^{\mathrm{T}}(\theta - \Gamma X) = \Gamma^{\mathrm{T}}R. \tag{7}$$

The algorithm starts with an initial guess for the solution, X_o, and builds an initial residual, r_o and a search direction, p_o. At every iteration k, an improved estimate for the residual r_k, the search direction p_k and the solution, X_k, is found. The minimization is stopped when the square of the residual, $r^{\mathrm{T}}r$, is below a given value, ϵ. The beauty of this algorithm is that the successive minimizations are carried out in a series of orthogonal conjugate directions. This means it is very fast, the solution can be found in typically 1 s of CPU time (running on a 1-GHz processor). As we shall see below, this is crucial to allow for the multiple minimizations required to estimate the accuracy of the method.

The minimization process has to be carried out through several iterations to arrive at a division of the lens plane into a grid that reflects well the uneven distribution of lensed images. For the first iteration we simply divide the lens plane into a regular grid. After this iteration, a first estimate of the mass is used to create a new grid (and a new Γ) where dense areas are sampled better than underdense areas.

The method has one potential pathological behaviour when applied to our problem. We cannot choose the minimization threshold, ϵ, to be arbitrarily small. If we choose a very low ϵ, the algorithm will try to find a solution that focuses the arcs in N_s sources, which are δ functions. This is not surprising as we are in fact assuming that all the $2N_\theta$ unknown β are reduced to just $2N_s$ β, i.e. the point source solution (see Paper I). Because the lensed galaxies are extended objects, such a solution is of course unphysical, and we therefore have to choose ϵ wisely. Because the algorithm will stop when $R^2 < \epsilon$, we should choose ϵ to be an estimate of the expected dispersion of the sources at the specified redshifts. This is the only prior which has to be given to the method. However, as shown in Paper I, the specific value of ϵ is not critical as long as it is within a factor of a few of the true source dispersion. As seen in Paper I, instead of defining ϵ in terms of R^2, it is better to define it in terms of the residual of the conjugate gradient algorithm, r_k^2. This speeds up the minimization process significantly:

$$\epsilon = r_k^{\mathrm{T}} r_k = R^{\mathrm{T}}\Gamma\Gamma^{\mathrm{T}}R. \tag{8}$$

As an example, 30 circular sources with a radius of 14 kpc located at redshifts between 1 and 6 typically correspond to $\epsilon = 2.0 \times 10^{-11}$.

2.2 Method accuracy

As seen above, it is crucial to stop the minimization before the absolute minimum of R^2 is reached. Because we are minimizing an N-dimensional quadratic function (R^2), the area where we stop is an N-dimensional ellipsoid around the global minimum. The end point of the minimization will then vary depending on the initial condition, X_o. That is, the solution is not unique as each minimization will stop in a different point on the N-ellipsoid. The physical meaning of this degeneracy is connected to our lack of knowledge about the shape of the sources. When traced back to the source plane, the pixels in the arcs are placed with any configuration within a compact region corresponding to the size of the source. This uncertainty in the shape of the sources can be accounted for by minimizing many times, each time with a different initial condition, X_o. Using a fast minimization algorithm such as the bi-conjugate gradient is therefore crucial in order to explore a large number of initial conditions and estimate the scatter in the final solution.

For the current analysis, the starting points for the minimization, X_o, are drawn from a uniform random distribution between 0 and $1.6 \times 10^{-3} \times 10^{15}\ h^{-1}\ \mathrm{M}_\odot$ for the masses and a random uniform distribution for the β positions in a box of 2 arcmin centred in the central dominant (cD) galaxy. The value 1.6×10^{-3} typically gives initial total masses of around $0.5 \times 10^{15}\ h^{-1}\ \mathrm{M}_\odot$ in the considered field of view.

There are also other factors which may reduce the accuracy of the method. One such source of uncertainty comes from the fact that the redshifts are not known with infinite precision but have a small uncertainty. For the majority of the lensed galaxies, the redshift has to be estimated using photometric data only, and errors of 15–20 per cent in redshift are quoted by B2005.

Inaccuracies in the redshifts are problematic for our reconstruction algorithm because they propagate into errors in the estimated angular diameter distances between us and the source, as well as between the source and the lens. These are of course crucial ingredients in calculating the Γ matrix for the linearized problem, and it is therefore important that we take this into consideration in our analysis.

To account for the redshift uncertainty, we again resort to multiple minimizations. We use different redshift realizations for the sources each time we solve for the lens equation (or equivalently minimize its quadratic residual, R^2). This allows us to propagate the error in the redshifts into scatter in the solution, and gives us an estimate of the inaccuracy of the solution through a frequentist approach. The redshifts are generated from a Gaussian distribution, with a mean and dispersion obtained from the data, which we assume is approximated by a Gaussian probability distribution for simplicity.

A final source of inaccuracy in the method is the adaptive gridding of the lens plane. As explained in Paper I, we take the initial grid to be regular and containing a low number of grid points. An 8×8 or 16×16 grid produces a nice initial solution, which looks roughly like a smooth version of the final solution. An adaptive grid is then created from this first solution. It is important that the maximum number of cells be chosen with caution. Too few cells may not sufficiently capture the details of the mass distributions. However, the number of grid cells should not be too high, exceeding the resolution set by the projected density of the observed images.

A natural upper limit for the number of cells is two times the number of pixels in the data (i.e. pixels forming part of one of the

arcs) minus two times the number of sources. A number of cells equal to this number would produce a square matrix Γ. For the analysis presented in this paper, we shall see in the next section that the number of pixels is $N_\theta = 601$. An estimate of the error due to choice of grid points can be obtained by repeating the analysis with different grid sizes.

It is important to emphasize here a key difference between the results presented in this work and those in Paper I. In that paper, many minimizations were performed to explore the space of solutions but with the difference that the grid was the same in all of them. That constant grid had ≈350 cells and was obtained after one minimization starting from a regular grid of 8×8 cells. Instead, in the present paper each minimization may have a different grid, which is obtained from a previous minimization. Allowing the grid to vary increases the dispersion of the solutions significantly. The interested reader will find that the standards set in Paper I are not reached in this paper. This is mainly a direct consequence of this extra source of variance in the modelling of the lens plane, which was not studied in Paper I.

3 ACS DATA

The data used in this paper are described in detail in B2005. Here we only briefly summarize their main characteristics. The original ACS image of A1689 (Benítez et al. 2002; B2005) was obtained after integration of 20 orbits with the *HST* in four bands (*G*, *R*, *I* and *Z*). The interested reader can find an overview of the ACS in Ford et al. (2003). The final published image covers a field of view of 3.3×3.3 arcmin^2 with a pixel size of 0.05×0.05 arcsec^2. The catalogue with the coordinates and redshifts of the arcs contains the positions and redshifts of 106 arcs, four of which have been spectroscopically identified in previous works (Fort et al., private communication; Frye et al., private communication). The bulk of the redshifts were estimated using the Bayesian software BPZ (Benítez 2000). In addition to the five bands mentioned above, the ACS observations were complemented with *U*-band observations obtained with the DuPont telescope at Las Campanas Observatory and *J*, *H*, *K* data at La Silla with the New Technology Telescope (NTT). With these bands, the final photometric redshifts are typically uncertain by 15–20 per cent (90 per cent confidence level). In the four cases where the redshifts are measured, the agreement with the photometric estimate is very good (less than 10 per cent). These errors, although large, play a weak role on the estimated mass basically because all the sources are beyond $z = 1$ where the angular diameter distance shows a much weaker dependency with the redshift than at $z < 1$. However, we will include the redshift uncertainty in our calculations.

The 106 arcs are associated with 30 systems or sources with redshifts in the range $1 < z < 6$. The positions in the catalogue correspond to the centre of the arc. We only use these central positions to identify the arcs. Then we carefully select all the pixels in each arc to build the final strong lensing data set. We go through all the tabulated positions and select the pixels belonging to the specified arc by eye. We only select the pixels which are clearly connected with the arc. In the cases where the arc is too faint, a smoothed version of the data is used to enhance the signal-to-noise (S/N) ratio. Eye selection is superior to algorithm selection in our case because software cannot be trusted to separate the faintest arcs from the background. After all the positions in the arcs have been selected, we repixelize the data in an area of 5×5 arcmin^2 using 512×512 pixel. Under this pixelization, the total number of pixels in our data set containing part of an arc is $N_\theta = 601$. The resulting data set is show in Fig. 1. These are the 601 θ positions which are used to invert the lens. The results are described in the next section.

Figure 1. Data used in the mass reconstruction. There are 106 arcs in this image, which are assigned to 30 different sources. Every arc has a flag associated to the putative source. Source redshifts range from $z \approx 1$ to $z \approx 6$. The area in this plot is similar to the field of view of the original data and it covers 3.3×3.3 arcmin^2.

As noted in the previous section, there is another important difference between the data set used in this work and the simulated data used in Paper I. The fraction of pixels forming part of a radial arc is significantly smaller in the former than in the latter case. This will have consequences in the recovered mass because the radial arcs are more sensitive to the very central regions in the mass distribution than the tangential arcs.

4 RECOVERED MASS DISTRIBUTION OF A1689

In this section we present the results of our analysis when applying the method of Section 2 on the data from Section 3. We show the results of 1000 minimizations, where the initial mass distribution and source redshifts are randomly varied. The maximum number of mass cells is approximately 600.

The result of this minimization process is shown in Fig. 2 where we compare the average of the 1000 recovered solutions with the ACS optical image of A1689. Keeping in mind that no information about the luminosity is used, the first obvious conclusion from this plot is the existing correlation between the luminous and the dark matter. The peak of the mass distribution falls very close to the cD galaxy. There is also a clear correlation between the position of the subgroup to the right and a secondary peak in the mass distribution. The small subgroup at ≈30 arcsec to the south of the cD galaxy seems to be sitting close to the top of other overdensity.

The substructure within 1 arcmin of the centre of the cluster suggests that the cluster is not fully relaxed. Another possibility is that some of the substructure arises from projection rather than from substructure within the main cluster. However, the existing correlation between the recovered mass and the galaxies suggests that the

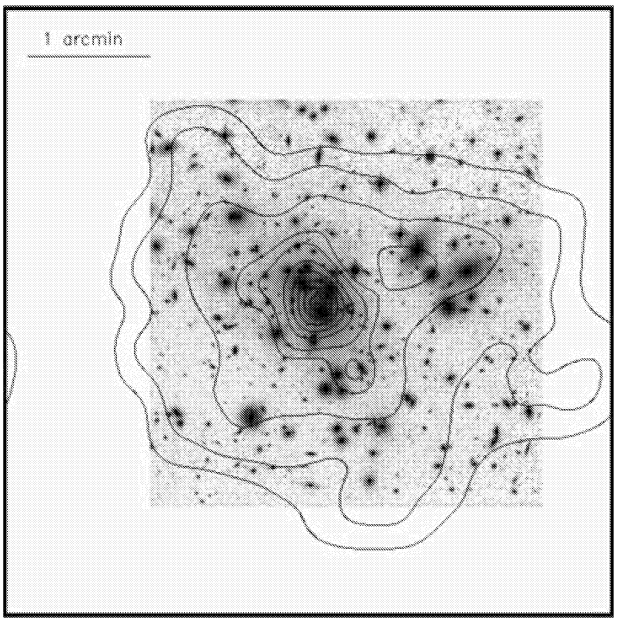

Figure 2. Mean recovered mass (contours) compared with the true ACS image. The mass is the average of 1000 minimizations of the lens equation where at each minimization we change the grid, the initial conditions, X_o and the redshifts of the sources. Contours go from 0.1 to 0.97 times the maximum mass density in intervals of 0.1 (0.097 last interval). Total mass in the field of view is about $5.2 \times 10^{14}\ h^{-1}\ \mathrm{M}_\odot$. The field of view in this plot and the others is 5×5 arcmin^2 unless otherwise noted.

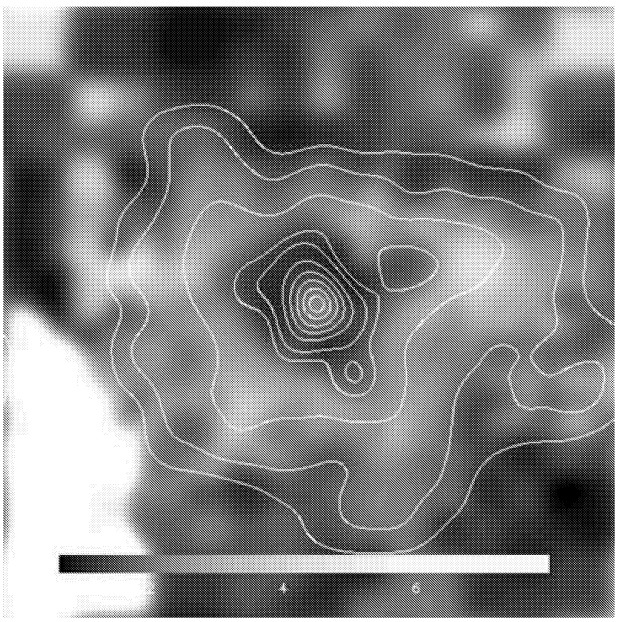

Figure 3. This grey-scale map shows the S/N ratio of the recovered mass which is obtained by dividing the mean recovered mass by the dispersion of the 1000 recovered maps. For clarity, the areas with S/N > 8 have been saturated (white colour). Note the low S/N ratio at about 20 arcsec from the centre of mass. The contours show the mean recovered mass of Fig. 2. The field of view is 5×5 arcmin^2.

substructure may be really present in the cluster. Another interesting feature is that the reconstructed mass seems to be insensitive to the external structure of A1689. There seems to be no significant structure beyond 2 arcmin from the cD galaxy. This can be explained if the mass distribution beyond this radius can be approximated by a spherical distribution. In this case, the Gauss theorem implies that the strong lensing data should be independent of the unknown outer mass distribution. If the distribution is highly elliptical, the recovered mass provided by SLAP may suffer additional biases. This is an interesting issue which we defer to a future investigation.

Looking at the dispersion of the 1000 minimizations tells us something about the reliability of our recovered mass profile. An estimate of the dispersion of these solutions can be seen in Fig. 3 where we plot the S/N ratio, which is defined as the ratio of the mean recovered map divided by the standard deviation map of the solutions. The first thing we should notice is that around 20 arcsec, the S/N ratio drops below 3. In other words, the mass estimate in this region cannot be trusted as well as in other regions. A similar behaviour can be observed at large radius as discussed above and may imply a degeneracy set by the limitations of the data we are using. The insensitivity of the data to the outer regions of the mass distribution is suggested also when we look at the average one-dimensional (1D) profile. The 1D density profile is defined as the average profile at a given distance from the centre normalized by the critical density, defined as

$$\Sigma_{\mathrm{crit}} = \frac{c^2}{4\pi G}\frac{D_\mathrm{s}}{D_\mathrm{d} D_\mathrm{ds}} = 4.29 \times 10^{15} \frac{h\mathrm{M}_\odot}{\mathrm{Mpc}^2}. \quad (9)$$

Here, we have assumed that Σ_{crit} is defined at the mean redshift of the sources, i.e. $z = 3$ (B2005). Note that the units of Σ_{crit} are $h\ \mathrm{M}_\odot\ \mathrm{Mpc}^{-2}$. These are the same as the recovered Σ, which is defined as

$$\Sigma = \frac{\mathrm{Mass}}{\mathrm{pixel}} = \frac{h^{-1}\,\mathrm{M}_\odot}{(h^{-1}\,\mathrm{Mpc})^2} = \frac{h\mathrm{M}_\odot}{\mathrm{Mpc}^2}. \quad (10)$$

The recovered 1D profile is shown in Fig. 4. Also shown is the dispersion of the 1000 recovered profiles. The dot-dashed line shows the best-fitting NFW profile (Navarro, Frenk & White 1995) found by B2005 using the same data. By comparing the reconstructed profile with an NFW profile we can confirm the excess found in B2005. This excess may also be well described by an NFW profile. We will

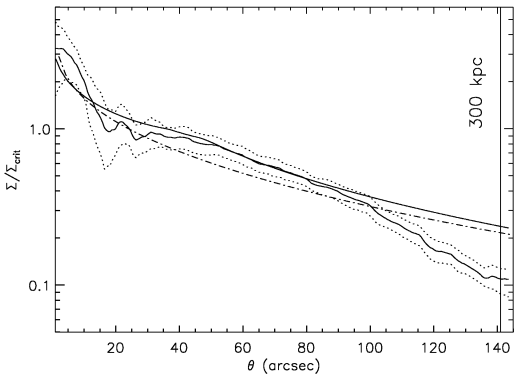

Figure 4. The plot shows the mean value (thin solid line) and the 95 per cent confidence region (dotted lines) of the 1D profiles for the 1000 minimizations in case (i). The dot-dashed line is the best-fitting NFW profile found in B2005. The density has been rescaled by the critical density, Σ_{crit}. The thick solid line is a very similar NFW profile plus an excess given by three NFW subhaloes around the main halo. See text for details.

discuss this point later but here we anticipate that the data are more likely to be compatible with an NFW profile plus an excess (thick solid line). However, it is clear from Fig. 4 that our reconstructed profiles differ significantly from an NFW profile at large radii, thus suggesting a possible bias in our results here. This possibility will be explored in more detail later. When we look at the normalized 1D profile (Fig. 4), we find another striking feature which also suggests possible bias in our results, this time in the very central region. As opposed to previous results based on the same data (B2005), the central density deviates from an NFW profile and even shows a dip at distances around 20 arcsec from the central peak. The same dip can be observed if we look at the map of the S/N ratio (see Fig. 3). This may be an indication that our algorithm is more sensitive to tangential than radial arcs. The radial images contain more information about the matter distribution in the very centre of the cluster than the tangential ones. This could be explained because we are minimizing the residual of the lens equation. The residual is basically dominated by the tangential arcs as they have more pixels than the radial arcs and therefore contribute more to the residual. Again, this possible bias will be explored later.

Finally, an interesting conclusion from Fig. 4 is that using a non-parametric algorithm does not mean necessarily that the solution cannot be well constrained within the error bars. In fact, these error bars are comparable to those obtained with parametric methods.

5 PREDICTED POSITIONS OF THE SOURCES

The solution found in the previous section also gives us the original position of the sources. Let us recall that in our algorithm we assume that the sources are point-like and they are described by just two numbers, namely the x and y coordinates at the centre. For each of the 1000 minimizations we obtain an estimate of the (x, y) position of each source. The result is plotted in Fig. 5. The recovered sources fall in a small area of $\approx 1 \times 1$ arcmin^2. Some sources seem to fall on top of others. Given the uncertainties in the photometric redshifts, it could happen that some of the sources are at the same redshift. Together with the fact that they appear in the same area in the sky, this makes us think that some of these sources may be the same. We should note, however, that previous work has identified a systematic problem when minimizing the lens equation in the source plane, namely the fact that the minimization is biased toward higher masses for the lens and with the sources being in a more compact region. If we are indeed affected by this, this would explain why the sources seem to fall in such a compact region. This possible systematic effect will be also studied later.

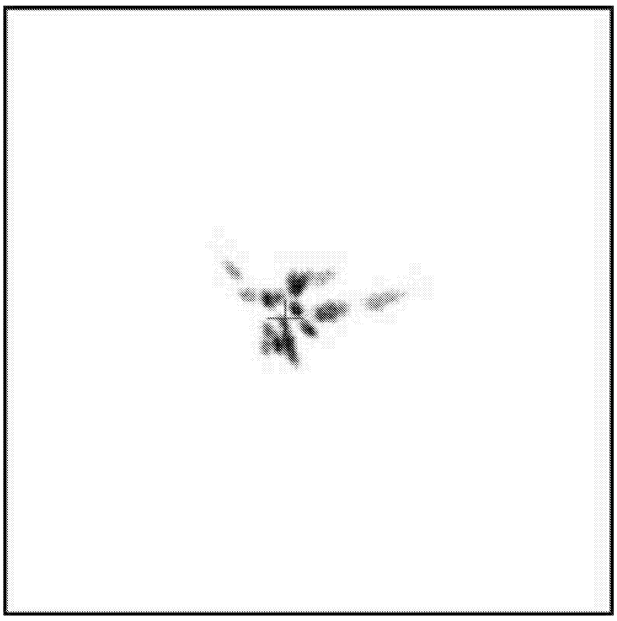

Figure 5. Zoomed version of the recovered β positions after 1000 minimizations. The field of view is 3.3×3.3 arcmin^2. The cross marks the position of the cD galaxy. Note how the small area of the source plane relative to the image plane and implies a high magnification of the background galaxies with a mean value of ~ 8.

6 CRITICAL CURVES

It is interesting to look at the critical curves of our reconstructed mass. These curves are defined as the regions where the magnification diverges. Normally one expects to see two types of curve: the tangential critical curve and the radial critical curve. The first is normally associated with the Einstein radius and is where the big tangential arcs tend to appear.

The radial critical curve defines the region where two multiple images merge or split in the radial direction. This curve is very interesting because it is sensitive to the particular profile of the inner region of the cluster. If we change the total mass, the concentration parameter and the characteristic scale, r_s, such that the tangential critical curve does not change much (i.e. we do not change the mass embedded within the giant tangential arcs) then we observe that smaller r_s produce smaller radial critical curves. In other words, the ratio between the tangential and the radial critical curves tells us something about the steepness of the profile between the radii of the giant arcs and the centre. A steep profile will produce a small relatively small radial critical curve, for a fixed tangential critical curve. A relatively large radial critical curve is generated by a flatter profile near the centre of the cluster. Note that for profiles steeper than the isothermal case, the radial critical curve is reduced to a point at the position of the lens.

Previous analysis of A1689 based on the same data (B2005) found a relatively large radial critical curve extending up to 20 arcsec from the centre of the cluster. NFW profiles are compatible with these large radial critical curves only if the halo characteristic radius, r_s, is relatively large. B2005 found best-fitting values of $r_s = 310\,h^{-1}$ kpc and concentration parameter $C_N = 8.2$ (with $C_N = R_{\rm virial}/r_s$). An NFW profile like this one reproduces well the derived critical curves in B2005.

The critical curves of our mean recovered model (see Fig. 2) are shown in Fig. 6. By comparing with the critical curves in B2005, we see that the inner curve (radial critical curve) is similar (or even larger in some areas) than that obtained in B2005. This fact suggests that the characteristic scale, r_s, must be indeed large, of the order of $300\,h^{-1}$ kpc or more. Also from the same plot, our critical curves show a smoother behaviour than previous analysis (B2005), which may suggest that we are not very sensitive to small details in the mass distribution. More specifically, the differences between our recovered critical curves and those found in B2005 are bigger in the case of the radial critical curve, which is more sensitive to the details in the central part of the cluster. A higher resolution is expected in the centre for the modelling of B2005 because the masses of the tight clump of luminous cluster galaxies found there are included in the model as part of the cluster sequence component (B2005). This level of detail is not easy to reproduce in detail with our non-parametric model, which would require more constraints in the centre for a more detailed fit here; hence, our results in the centre $r < 20$ arcsec

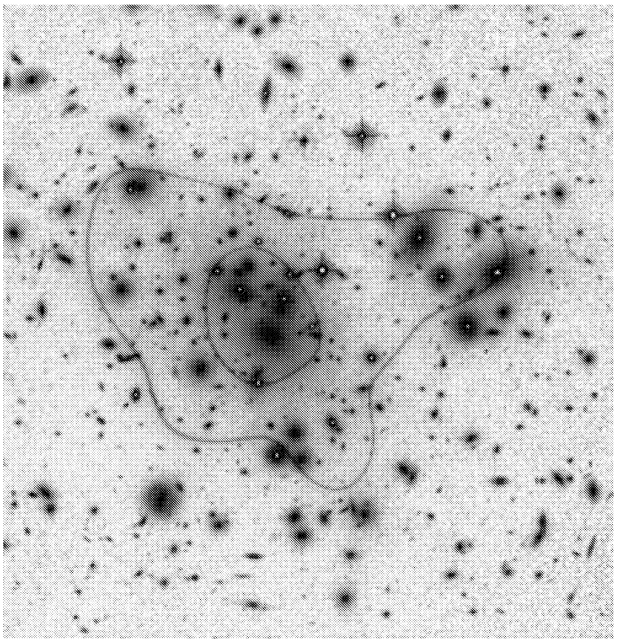

Figure 6. Critical curves for the mean recovered mass in Fig. 2. The field of view is the same as the original image (3.3 arcmin). Note the clear formation of a radial critical curve whose size relative to the tangential critical curve requires a shallow central mass profile.

should probably be regarded as a somewhat smoothed version of the central mass profile. This very last point may be connected with the drop interior to the critical curve (around 20 arcsec from the centre) in the mass density profile (see Fig. 4). This feature in the profile could be due to a degeneracy among the masses in the cells in the very central region of the cluster and could be easily explained by the argument used above that our algorithm is less sensitive to the radial than to the tangential arcs. This is also made evident when we realize that the high S/N regions in Fig. 3 coincide with the tangential critical curve, while the radial critical curve falls in a region with low S/N ratio. The features in the profile may be real or due to a systematic bias in our algorithm. Answering this question is the purpose of the next sections.

7 ERROR ANALYSIS AND POSSIBLE SYSTEMATICS

The results in the previous section offer some answers about the mass distribution in A1689 but also raise some serious questions about the reliability of our results. A visual comparison with the results of B2005 indicates some disagreement between our mass distribution and theirs. Our recovered mass distribution shows substructure within the central $200\,h^{-1}$ kpc (1.5 arcmin) with dips and peaks around the central peak. The overall mass distribution is similar in shape to that of B2005, but with more pronounced substructure. The difference can be partially explained by the fact that parametric methods implicitly assume a smooth distribution for the main dark matter component with no dips, while we do not. The second possibility is that the dips are an artefact coming from degeneracies of the modelling procedure. As shown in Paper I, we may expect a variety of models to be consistent with the data. Some of these models may show degeneracies between neighbouring cells at small scales if the result is not sensitive to these small scales, although in general is not possible to predict where the degeneracies will appear. We expect that the range of valid models reduces as the number of arcs increases. This means that each case has to be studied separately. This possibility will be explored further in the next section.

In this section we focus on another source of systematics. In Section 4 we included in our analysis the numerical uncertainties in our algorithm. These were the uncertainty in the knowledge of the redshift of the sources and the uncertainty in the shape of the sources. The uncertainty in the redshift was included by assigning different redshifts to the sources at each minimization (Gaussian distribution), while the uncertainty in the shape of the sources was included by minimizing many times, each one with a different initial condition, X_o.

In Section 4 we also changed the grid at each iteration using our dynamical adaptive grid, which constructs the new grid based on the previous solution. For doing this, we had to fix one parameter of the algorithm, the total number of cells, N_c. The algorithm needs another parameter to be defined, namely the minimum residual we want to achieve, ϵ. The algorithm stops when $R^2 < \epsilon$, where ϵ can be defined by the size of the sources and their number. In Sections 2 and 4 we gave some intuitive motivation on how to choose ϵ and N_c, respectively. In this section we address the issue of how sensitive the results are to these two parameters.

We consider three different scenarios or cases.

(i) The minimization is performed with a number of cells $N_c \approx 600$ and $\epsilon = 2 \times 10^{-11}$. This is the case used to present the results in Section 4.

(ii) As in case (i), but we reduce the number of cells to $N_c \approx 300$.

(iii) As in case (i), but we reduce the size of the sources to $\epsilon = 5 \times 10^{-12}$.

Case (i) was already studied in the previous sections and is used here for comparison. For each of the cases (ii) and (iii), we run another 1000 minimizations changing the starting point, X_o, the redshifts and the grid as we did in case (i) (Section 4).

In case (ii), by reducing the number of cells we reduce the number of possible solutions, i.e. we reduce the uncertainty in the solution. We also degrade the resolution because we have to fill the same space (5×5 arcmin^2) with half the number of cells. After averaging 1000 minimizations, the recovered mass distribution[1] looks similar to that found in case (i), with the main difference being in the outer areas where case (ii) shows an even larger deficit in mass when compared to the NFW profile. The critical curves[2] also look very similar to those found in case (i) but showing a slightly larger radial critical curve, which suggests a higher concentration of mass near the centre of the cluster. The average of the 1D profiles together with its 95 per cent error bars can be seen in Fig. 7. The plot clearly demonstrates the departure from the NFW profile at large radii. It also shows the reduction in the dispersion of the solutions as well as a lack of a dip at 20 arcsec. The same effect can be seen when we look at the predicted position of the sources (Fig. 8). Contrary to what happened in case (i) (see Fig. 5), the predicted positions of the sources in case (ii) do not suggest a smaller number of sources. A closer look reveals that in case (ii) the smaller number of cells produces a sequence of grids with very small differences between them. In other words, in case (ii) we are in a situation in which the

[1] Figure available at http://darwin.physics.upenn.edu/SLAP/ and as Supplementary Material to the online version of this article.

[2] Figure available at http://darwin.physics.upenn.edu/SLAP/ and as Supplementary Material to the online version of this article.

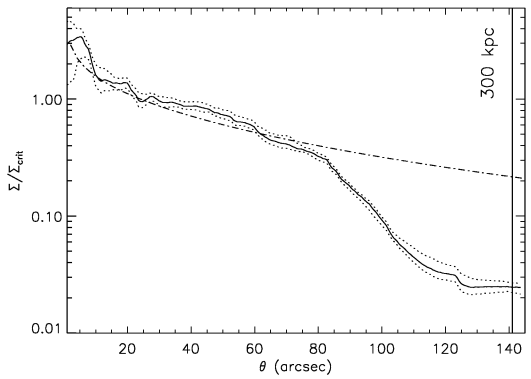

Figure 7. Recovered 1D profile (solid line) with error at 95 per cent level (dotted lines). The dot-dashed line is the best NFW profile found in B2005. This is for case (ii) (300 cells, 2×10^{-11}).

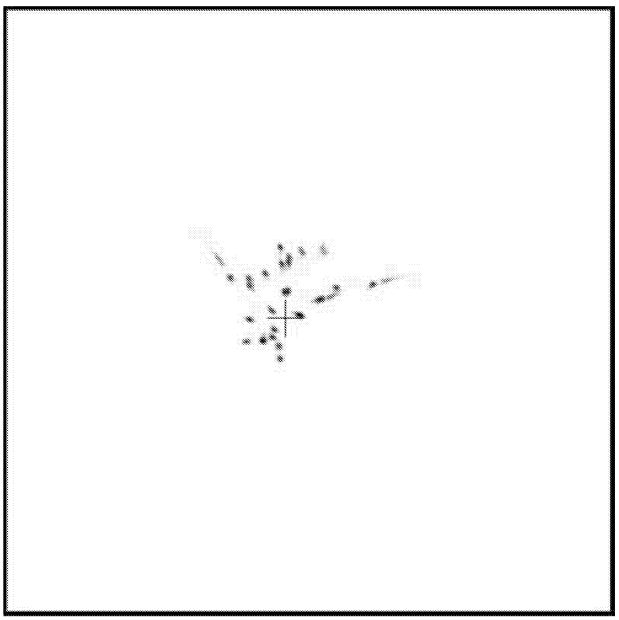

Figure 8. Zoomed version of the recovered β positions after 1000 minimizations for case (ii). Field of view is 3.3×3.3 arcmin^2.

grid has been practically fixed from iteration 1. This fact contributes crucially to the reduction in the range of solutions (masses and β positions).

Case (iii) is interesting to explore because it forces the algorithm to find a solution closer to the unphysical point source solution. The total dispersion in the source plane has now been reduced by a factor of 4. The solutions achieve this by adding more substructure to the mass distribution, and when ϵ is made small enough, the β positions are also shifted toward the position of the centre of mass. This effect is well known and it was studied in Paper I. In our particular case, the mean mass distribution of the 1000 solutions looks again similar[3] to that found in case (i) but showing more substructure. The average

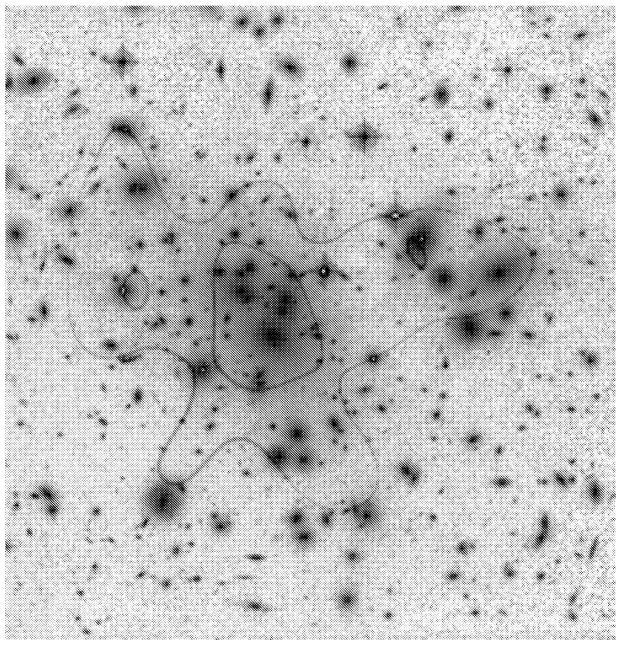

Figure 9. Critical curves for the mean recovered mass in the case (iii). The field of view is (3.3 arcmin)^2.

1D profile[4] is also similar to that in Fig. 5. Here we present only the critical curves in Fig. 9 where the effect of the extra substructure can be appreciated.

The residual, R or ϵ, has the physical meaning of being the variance or size of the sources. Setting a very small ϵ produces a biased mass distribution, which focuses the arcs into very small sources or point sources. The point source solution achieving this is normally unphysical, as was shown in Paper I. On the contrary, choosing a large ϵ will stop the minimization early, resulting in a short-sighted cluster, meaning the solution cannot focus the arcs properly. This short-sighted cluster solution is normally a smoother, lower-mass version of the real solution.

8 TESTING THE RESULTS WITH SIMULATIONS

The previous section has two possible interpretations. On the pessimistic side, we have raised concerns about the reliability of our results because we show how the results change depending on our choice of number of cells and the stopping point of the minimization. On the other hand, the positive interpretation is that the change in the results is not dramatic and our conclusions seem to be relatively insensitive to big changes in the minimization process.

Although the last section gave us an idea about the dispersion in the solution, it did not address the issue of whether or not the recovered solution is biased. The problem in answering this question is, of course, that we do not know what the real mass distribution is, and thus there is nothing to compare our results with. The aim of this section is to rectify this by using a simulated data set which mimics the main features of the real data. With a simulation we can easily check aspects such as how sensitive the data are to the mass

[3] Figure available at http://darwin.physics.upenn.edu/SLAP/ and as Supplementary Material to the online version of this article.

[4] Figure available at http://darwin.physics.upenn.edu/SLAP/ and as Supplementary Material to the online version of this article.

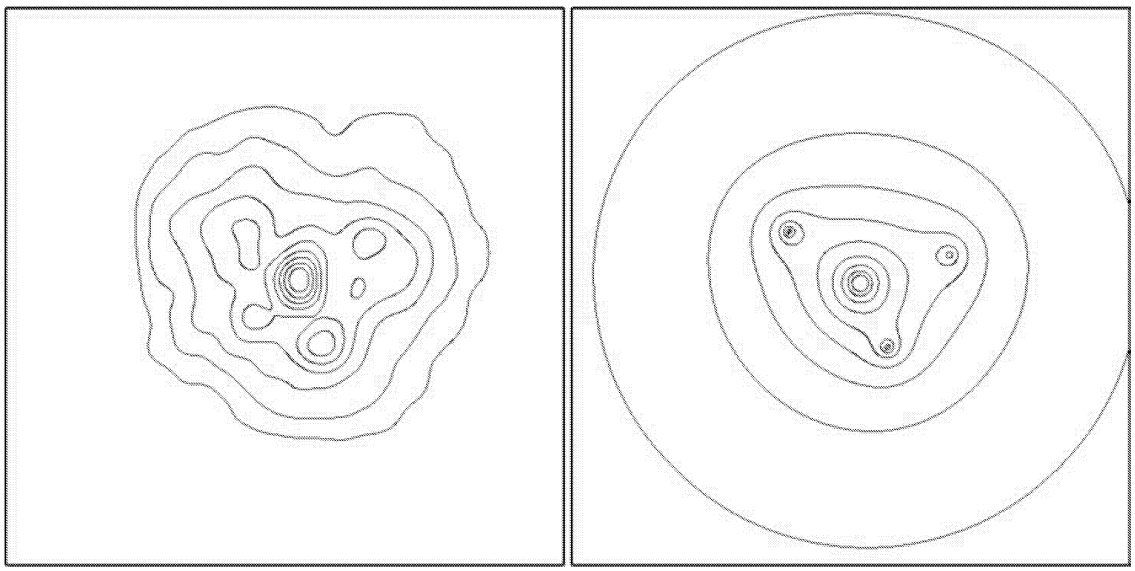

Figure 10. Recovered mass (left) compared with the original simulated one (right). The contours increase in steps of 0.1 times the maximum central density (in the recovered case) starting at 0.1 times the maximum. The field of view is (5 arcmin)^2.

distribution in the very centre or in the area beyond the tangential arcs. Our simulated cluster is a simplified version of the recovered mass distribution, made up of a superposition of NFW profiles. Because the recovered solution has a mass deficit in the outer parts, the simulated cluster has a larger total mass, but is chosen so that it resembles well the mass distribution within the giant arcs.

We use a total mass of $M_{\rm T} = 0.68 \times 10^{15}\ h^{-1}\ {\rm M}_{\odot}$ in the field of view (5×5 arcmin^2). For simplicity, our simulated cluster is made of only four NFW haloes. The main halo is assigned a mass of $M_1 = 0.53 \times 10^{15}\ h^{-1}\ {\rm M}_{\odot}$ and placed at the maximum of the averaged recovered mass in Section 4. The second halo is given a mass of $M_2 = 0.07 \times 10^{15}\ h^{-1}\ {\rm M}_{\odot}$ and centred in the north-eastern subgroup. The third halo with $M_3 = 0.03 \times 10^{15}\ h^{-1}\ {\rm M}_{\odot}$ is centred to the south-east of the main group, and finally the fourth halo with $M_4 = 0.05 \times 10^{15}\ h^{-1}\ {\rm M}_{\odot}$ is placed to the north-west of the main halo (see Fig. 10). This simulated cluster resembles the reconstructed mass profile found in Section 4 but with the difference that it has a sharp cusp in the centre (plus three off-peak sharp cusps) and the tails of the distribution do not fall off as quickly as in the recovered mass distribution. We have also verified that the model roughly reproduces the recovered critical curves.[5] The 1D profile of this simulated cluster is shown in Fig. 4 (thick solid line) where it is compared with the reconstructed 1D profile and the best-fitting NFW profile of B2005. The projected mass distribution is shown in Fig. 10. For the lensing simulation, the cluster is located at the same redshift as A1689 ($z = 0.18$).

The second ingredient of the simulation is the sources. We use 30 sources whose β positions are taken as random within a box of 1×1 arcmin^2 around the centre of the main halo. The sources are assumed to be circular with radii of a few kpc, and are placed at the redshifts published in B2005.

The final part of the simulation is to find the arcs corresponding to the previous configuration.[6] For this we use a simple ray-tracing algorithm. With these simulated data we follow a reconstruction process similar to case (i) in the previous section. We run 1000 minimizations (but with 500 cells instead of 600 and with $\epsilon = 2.0 \times 10^{-11}$) and calculate the mean value and dispersion of the solutions.

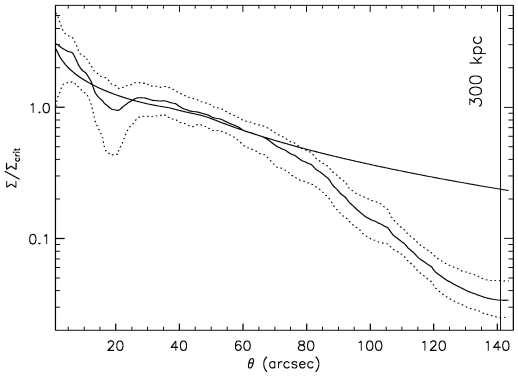

Figure 11. The mean and error region (95 per cent level) of the 1000 recovered profiles after changing the initial conditions, $X_{\rm o}$, the redshifts and the grid at each minimization. The thick solid line is the original profile of the simulated cluster.

The average of the 1000 recovered masses is shown in Fig. 10 (left) and it is compared with the original mass distribution (right). The position of the main halo is reconstructed with good accuracy. In the position of the secondary haloes we reproduce an overdensity, although a spurious overdensity also appears in the south-west of the main halo. The total recovered mass is $4 \times 10^{15}\ h^{-1}\ {\rm M}_{\odot}$, that is 40 per cent smaller than the original total mass. This deficit in mass is again concentrated in the outer areas, beyond the position of the giant arcs, as can be seen from the recovered 1D profile (Fig. 11). The simulation confirms that the algorithm is insensitive to the mass distribution beyond the most distant arcs from the centre. It is important to note that the bias in the recovered profile in the outer regions is not due to any intrinsic prior in the masses. The

[5] Figure available at http://darwin.physics.upenn.edu/SLAP/ and as Supplementary Material to the online version of this article.
[6] Figure available at http://darwin.physics.upenn.edu/SLAP/ and as Supplementary Material to the online version of this article.

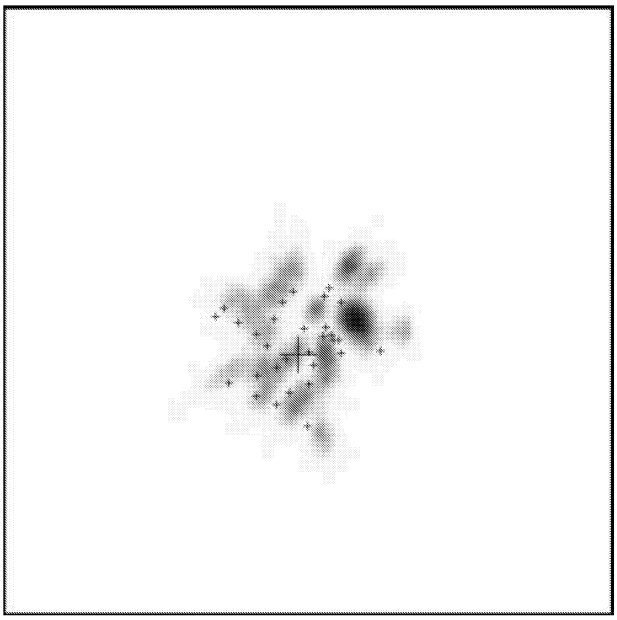

Figure 12. Reconstructed positions of the sources (grey-scale) for 1000 minimizations. The true position of the sources is marked with small crosses. The big cross is the position of the main halo. The field of view is a zoomed version (2×2 arcmin^2) of the original 5-arcmin field of view.

recovered profile in the outer regions can be biased toward both lower or higher values when compared with the true profile. In fact, the recovered profile in these regions remains close to the profile given by the initial condition. Starting with low masses in X_o (as we do in our analysis) will produce a profile with low masses in the tails. On the contrary, if X_o has high masses in the outer regions, these will remain high at the position where the minimization is stopped. This is just a manifestation of the low sensitivity of the outer regions to the data. The small error bars in the outer regions can also be explained by the fact that the solution retains a memory of the initial conditions in these regions and also that the initial densities (mass/cell) are systematically smaller in these regions due to the griddification of the lens plane (big cells in the outer regions). On the other hand, the recovered profile in the central region is much less sensitive to the initial conditions, producing in all cases profiles which more or less resemble the original one. However, we observed that starting with small initial conditions renders better results in the central regions than using large initial conditions. This fact can be connected with the point source solution where large fluctuations are needed to achieve convergence. Another interesting conclusion from Fig. 11 is that the algorithm also seems to have some problems finding the right mass in the central region. It overpredicts the central density and underpredicts the density in the area near the radial critical curve. It even suggests a fictitious dip in this area. When we repeat the same exercise but reducing the number of cells down to 300 (and keeping $\epsilon = 2 \times 10^{-11}$), we observe a similar behaviour to that described in Section 7.[7] The recovered 1D profile does not show a dip at 20 arcsec and the profile falls faster at radii larger than 60 arcsec. Between 20 and 60 arcsec, the 1D profile overpredicts the real one by about 20 per cent in the case with 300 cells.

An interesting lesson can be learned when we combine both results. The recovered mass distribution interior to the radial critical curve is closer to the real one when we use a smaller number of cells (300); however, between the radial and the tangential critical curves, the recovered mass profile is better when we increase the number of cells (500–600). Unsurprisingly, we are also able to conclude that we definitely recover a biased mass distribution beyond 70 or 80 arcsec from the centre.

Regarding the location of the sources, the recovered β positions deviate from the true position by between 0 and 5 arcsec (see Fig. 12). Reducing the number of cells from 500 to 300 does not show any appreciable improvement in this situation and the recovered β positions look almost indistinguishable from Fig. 12. This is to be contrasted by case (ii) in Section 7. However, as opposed to that case, reducing the number of cells to 300 in the simulated data does not here produce a sequence of almost identical grids. This suggests that the recovered positions of case (ii) in Section 7 (see Fig. 8) are more the product of fixing the grid than being the real position of the sources.

9 CONCLUSIONS

Using a non-parametric algorithm (SLAP) we reconstruct the mass distribution of A1689 based on strong lensing data containing the 106 multiply lensed images identified by B2005. The reconstructed mass agrees well with previous estimations based on parametric algorithms (B2005). Our non-parametric approach is an essential complement to the more model-dependent methods and also allows us to understand better the uncertainties and potential ambiguities involved in using strong lensing data for generating surface mass distributions. In particular, we find that our recovered mass is biased toward smaller values beyond the most external tangential arcs and there is some evidence for degeneracy problems in the very central region. However, we also conclude that the total mass can be well constrained within 70 arcsec from the centre of the cluster. The total projected mass within 70 arcsec from the centre is found to be $0.25 \times 10^{15}\ h^{-1}\ \mathrm{M}_{\odot}$. The simulated work suggest that the estimated profile between 20 and 70 arcsec is reliable. Using this profile we can try to constrain its slope. In Fig. 13 we show two examples of power laws. The case $n = 2$ corresponds to the isothermal sphere. We also show the modified power laws when we add a core of $15\,h^{-1}$ kpc.

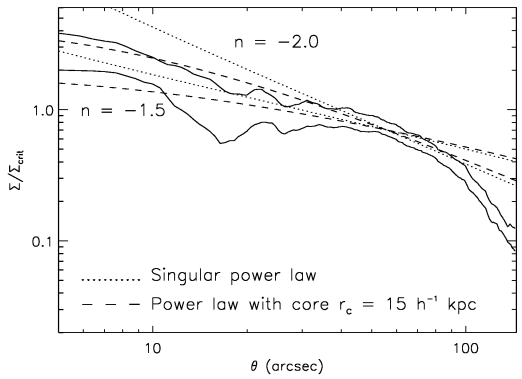

Figure 13. Power laws compared with the recovered mass profiles (the confidence region is taken from Fig. 4). We show two cases, the singular power law $\rho(r) \propto r^{-n}$ (dotted line) and the power law with core $\rho(r) \propto (r + r_c)^{-n}$ (dashed line).

[7] Figure available at http://darwin.physics.upenn.edu/SLAP/ and as Supplementary Material to the online version of this article.

The singular isothermal sphere (singular power law with $n = 2$) is ruled out by our data but it may be compatible when we add a core of about $15\,h^{-1}$ kpc. It is interesting to compare this plot with the results of Broadhurst et al. (2005a) where they find a best fit for the slope of $n = 1.65$. Also using simulations we show how the degeneracy in the central region can be reduced by taking a smaller number of cells, which naturally decreases the degrees of freedom. This is done at the expense of a bias in the outer regions, which is increased when the number of cells is low. Testing the algorithm with simulations which mimic the real data and the average estimated mass, we found that the best results can be obtained combining a minimization with a relatively large number of cells ($N_c \approx 500$) with a minimization with a smaller number of cells ($N_c \approx 300$). Combining these results we find that the mass recovered in a non-parametric way is compatible with an NFW profile plus an excess associated to substructure around the central overdensity.

Although not discussed in this work, it is worth mentioning that the algorithm fails to predict the redshift for other arcs present in the ACS data and with no current estimate of their redshifts. Most of the tested new arcs show a flat probability density function beyond redshift $z = 1$. The situation can be improved if pairs of arcs are identified in the data.

Our modelling indicates that the central region of the cluster is either affected by projection along the line of sight or is not yet fully relaxed as significant local density perturbations are found in our reconstruction. Evidence of ongoing merging has been also reported from an analysis of recent X-ray data (Andersson & Madejski 2004). The mass derived from the X-ray profile differs from that derived here when the cluster is assumed to be in a relaxed state (Andersson & Madejski 2004). If we believe the lensing results, it means the assumption of hydrostatic equilibrium used to derive the mass from X-rays may be hard to justify in detail (Xue & Wu 2002).

Previous analyses of A1689 using different lensing techniques support this hypothesis as they tend to agree in the mass. Our integrated mass estimate agrees well with these previous analyses.[8]

Tyson & Fischer (1995) found a mass of $(0.18 \pm 0.01) \times 10^{15}\ h^{-1}\ \mathrm{M}_\odot$ enclosed in a radius of 51 arcsec from the centre. In the same radius we find a mass of $(0.16 \pm 0.02) \times 10^{15}\ h^{-1}\ \mathrm{M}_\odot$. At larger radii Taylor et al. (1998) found $(0.5 \pm 0.09) \times 10^{15}\ h^{-1}\ \mathrm{M}_\odot$ within 108 arcsec, Dye et al. (2001) found $(0.48 \pm 0.16) \times 10^{15}\ h^{-1}\ \mathrm{M}_\odot$ within 112 arcsec, while we found $(0.42 \pm 0.07) \times 10^{15}\ h^{-1}\ \mathrm{M}_\odot$ in 110 arcsec. At radii larger than 70 arcsec we believe our total mass is biased toward lower values due to the insensitivity of the outer regions to the data.

In the literature we can find numerous studies of how masses derived from X-rays, optical and lensing compare (Miralda-Escudé & Babul 1995; Allen 1998; Wu et al. 1998; Wu 2000; Cypriano et al. 2004). Systematically, a discrepancy of about 2 is found in the central regions of some clusters, especially in those with evidence of being in a non-relaxed state (Allen 1998). A combination of the gravitational potential in the central region derived from strong lensing observations with high-resolution X-ray data will allow exciting studies focusing on the dynamical state of the gas in these regions. Also interesting is to combine the strong lensing results in the central region with weak lensing information, which allows us to extend the analysis up to Mpc scales (Bradac et al. 2005; Broadhurst et al. 2005b).

[8] Figure available at http://darwin.physics.upenn.edu/SLAP/ and as Supplementary Material to the online version of this article.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation CAREER grant AST-0134999, NASA grant NAG5-11099, the David and Lucile Packard Foundation and the Cottrell Foundation. The authors would like to thank Elizabeth E. Brait, M. Takada and E. Hayashi for useful discussions.

REFERENCES

Abdelsalam H. M., Saha P., Williams L. L. R., 1998a, MNRAS, 294, 734
Abdelsalam H. M., Saha P., Williams L. L. R., 1998b, AJ, 116, 1541
Allen S. W., 1998, MNRAS, 296, 392
Andersson K. E., Madejski G. M., 2004, ApJ, 607, 190
Benítez N., 2000, ApJ, 536, 571
Benítez N. et al., 2002, Bull. Am. Astron. Soc., 34, 1236
Bradac M., Schneider P., Lombardi M., Erben T., 2005, A&A, 437, 39
Broadhurst T., Huang X., Frye B., Ellis R., 2000, ApJ, 534, L15
Broadhurst T., Benítez N., Coe D., Sharon K., Zekser K., White R., Ford H., Bouwens R., 2005a, ApJ, 619, 143 (B2005)
Broadhurst T., Takada M., Umetsu K., Kong X., Arimoto N., Chiba M., Futamase T., 2005b, ApJ, 619, L143
Cypriano E. S., Sodré L. Jr, Kneib J. P., Campusano L. E., 2004, ApJ, 613, 95
Diego J. M., Protopapas P., Sandvik H. B., Tegmark M., 2005, MNRAS, 360, 477 (Paper I)
Dye S., Taylor A. N., Thommes E. M., Meienheimer K., Wolf C., Peacock J. A., 2001, MNRAS, 321, 685
Ford H. C. et al., 2003, in Blades J. C., Siegmund O. H. W., eds, Future EUV/UV and Visible Space Astrophysics Missions and Instrumentation. Proc. SPIE, 4854, 81
Gavazzi R., Fort B., Mellier Y., Pello R., 2003, A&A, 403, 11
Kneib J.-P., Mellier Y., Fort B., Mathez G., 1993, A&A, 273, 367
Kneib J.-P., Mellier Y., Pello R., Miralda-Escudé J., Le Borgne J.-F., Boehringer H., Picat J.-P., 1995, A&A, 303, 27
Kneib J.-P., Ellis R. S., Smail I. R., Couch W., Sharples R., 1996, ApJ, 471, 643
Kochanek C. S., Schneider P., Wambsganss J., 2004, in Meylan G., Jetzer P., North P., eds, Part 2 of Gravitational Lensing: Strong, Weak and Micro, Proceedings of the 33rd Saas-Fee Advanced Course. Springer-Verlag, Berlin
Miralda-Escudé J., Babul A., 1995, ApJ, 449, 18
Navarro J. F., Frenk C. S., White S. D. M., 1995, MNRAS, 275, 720
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1997, Numerical Recipes in Fortran 77. Cambrdige Univ. Press, Cambridge
Saha P., Williams L. L. R., 1997, MNRAS, 292, 148
Saha P., Williams L. L. R., 2004, ApJ, 127, 2604
Sand D. J., Treu T., Ellis R. S., 2002, ApJ, 574, 129
Taylor A. N., Dye S., Broadhurst T. J., Benítez N., van Kampen E., 1998, ApJ, 501, 539
Treu T., Koopmans L. V. E., 2004, ApJ, 611, 739
Trotter C. S., Winn J. N., Hewitt J. N., 2000, ApJ, 535, 671
Tyson J. A., Fischer P., 1995, ApJ, 446, L55
Tyson J. A., Kochanski G. P., Dell'Antonio I., 1998, ApJ, 498, L107
Warren S. J., Dye S., 2003, ApJ, 590, 673
Williams L. L. R., Saha P., 2001, AJ, 119, 439
Wu X-P., 2000, MNRAS, 316, 299
Wu X-P., Chiueh T., Fang L-Z., Xue Y-J., 1998, MNRAS, 301, 861
Xue S-J., Wu X-P., 2002, ApJ, 576, 152

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article online.

Figure S1. Recovered mass for case (ii) in the paper. The field of view is 5 arcmin.

Figure S2. Critical curve for case (ii) in the paper. The field of view is 3.3 arcmin.

Figure S3. Recovered mass for case (iii) in the paper. The field of view is 5 arcmin.

Figure S4. 1D profile after 1000 minimizations for case (iii) in the paper. Error bars are 95 per cent. Dot-dashed line is the best NFW profile in Broadhurst et al. 2005.

Figure S5. Critical curves of the simulated cluster. The field of view is 5 arcmin.

Figure S6. Simulated arcs from the simulated cluster. Each source has a different colour. There are 30 sources randomly distributed in space and with redshifts between 1 and 6 producing about 80 arcs in this image.

Figure S7. Reconstructed profile with error bars (95 per cent) for the simulated cluster after 1000 minimizations and using 300 cells. The solid line is the original 1D profile of the simulation.

Figure S8. Solid line is our total integrated (projected) mass as a function of radius. The symbolds are previous estimates from Tyson & Fischer (star), Taylor et al. (triangle) and Dye et al. (square) using different lensing techniques.

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