SLAC-PUB-11563
November 2005

Atomic inference from weak gravitational lensing data

Phil Marshall
Kavli Institute for Particle Astrophysics and Cosmology, Stanford University PO Box 20450, MS29, Stanford, CA 94309, USA

ABSTRACT

We present a novel approach to reconstructing the projected mass distribution from the sparse and noisy weak gravitational lensing shear data. The reconstructions are regularised via the knowledge gained from numerical simulations of clusters, with trial mass distributions constructed from n NFW profile ellipsoidal components. The parameters of these "atoms" are distributed a priori as in the simulated clusters. Sampling the mass distributions from the atom parameter probability density function allows estimates of the properties of the mass distribution to be generated, with error bars. The appropriate number of atoms is inferred from the data itself via the Bayesian evidence, and is typically found to be small, reflecting the quality of the data. Ensemble average mass maps are found to be robust to the details of the noise realisation, and succeed in recovering the demonstration input mass distribution (from a realistic simulated cluster) over a wide range of scales. As an application of such a reliable mapping algorithm, we comment on the residuals of the reconstruction and the implications for predicting convergence and shear at specific points on the sky.

Key words: methods: gravitational lensing

1 INTRODUCTION

Mapping mass distributions in clusters via the weak gravitational lensing effect has become something of a standard tool in astrophysics, allowing these most massive objects to be better understood in terms of their matter content, dynamical state and their value as galaxy evolution laboratories and cosmic observtaories. Given this importance, it seems worthwhile to investigate more accurate, more robust, and more practically useful, methods for reconstructing the mass distributions in clusters from the available data.

Following a number of seminal papers on the subject in the 1990s (*e.g.* Tyson et al. 1990; Kaiser & Squires 1993), the emphasis now is very much on the application of mapping methods to weak gravitational lensing shear data. Large CCD mosaic cameras such as SuprimeCam at Subaru, MegaCam at CFHT and the ESO Wide Field Imager at La Silla have enabled the mass distributions of clusters to be mapped to much larger radii than before (*e.g.* Clowe & Schneider 2002; Broadhurst et al. 2005). From space with HST, the same science has been made possible in higher redshift clusters, first though large multi-pointing datasets (*e.g.* Hoekstra et al. 2000; Kneib et al. 2003) and with observations with the Advanced Camera for surveys (ACS) (*e.g.* Lombardi et al. 2005; Jee et al. 2005). As is always the case in astronomy, these data are being pushed to new limits: the aim now is to understand cluster mass distributions in great detail, moving beyond the simple mass estimates of the early years. Kneib et al. (2003) and Gavazzi et al. (2003) measured the outer logarithmic slope of the density profile in two systems, while Clowe et al. (2004) investigated the relative peak positions of the gravitating mass density and the intracluster gas density (the latter being derived from the X-ray surface brightness). The quantification of the substructure in galaxy clusters is a topic of ongoing research, with progress being made in the central parts of clusters by comparing strong and weak lensing mass models with predictions from N-body simulations (Natarajan & Springel 2004).

Despite the advances in data quality, weak gravitational lensing data remains very sparse and very noisy. It is notable indeed that the most exciting results in the field in recent years have come from the comparison of weak lensing data with external observations, such as the modelling of strong gravitational lensing features (*e.g.* Kneib et al. 2003; Bradač et al. 2005), the X-ray emission (as in Clowe et al. 2004), and the optical data on the cluster member galaxies (Czoske et al. 2002; Kneib et al. 2003). We take a cue from this observation as we ask how best to extract as much information as possible from the weak lensing data, and draw the most meaningful conclusions about the mass distributions in clusters. In general terms, including information from external sources in order to understand better a weak lensing cluster means assigning appropriate prior probability distribution functions to whatever set of parameters we are

Submitted to Monthly Notices of the Royal Astronomical Society

Work supported by Department of Energy contract DE-AC02-76SF00515

using to model the cluster: to this end we seek a flexible fitting algorithm that is able to cope with such constraints and return parameter estimates (with accurate uncertainties) that reflect all the observational data in hand. Moreover, the choice of model itself can be made so as to facilitate comparisons with analyses of independent observations. However, in the first instance, it seems sensible to postpone combination with other observational data until the lensing signal is understood.

In all weak lensing mass mapping algorithms to date this model has been that of a grid of mass map pixels, whose values comprise the model parameters, an approach often termed "parameter-free" reconstruction. It is not at all clear that a grid of pixels is the optimal model for a cluster mass distribution; as shown in (Marshall et al. 2002,hereafter M02), such a large number of parameters is very often discouraged by the data quality, leading to over-fitting and potentially over-interpretation of the data. In M02, the number of parameters was reduced by including the assumption of the mass pixels being correlated on some characteristic angular scale (typically ~ 1 arcminute), which led to smoother, less noisy maps, but which, by virtue of the resolution scale being inferred from the data themselves, were restricted to show only "believable" structures with angular scales greater than the resolution parameter. In this work we seek a more natural basis set of functions with which to model cluster mass distributions. By correlating pixels together, cluster-like structures can be more easily modelled: the logical extension of this idea is to build up a mass distribution from components that already have cluster-like properties.

From N-body simulations we expect the ensemble average mass distribution to be ellipsoidal with an NFW profile (*e.g.* Navarro et al. 1997; Jing & Suto 2002), and that clusters lie at the high mass end of a hierarchy of structures, each with this same universal profile. The NFW profile has some support from the data, at least for the most massive haloes: previous gravitational lensing analyses have found the NFW profile mass distribution to provide a somewhat better description of the data than competing models (*e.g.* Clowe & Schneider 2002; Kneib et al. 2003; Gavazzi et al. 2003), as have high resolution X-ray studies (*e.g.* Allen et al. 2002). If all the mass in clusters of galaxies were distributed exactly in elliptical NFW-profile haloes, then the optimal basis set for the lensing inverse problem might be expected to be a group of elliptical NFW profile mass components, shifted and scaled to match the various mass clumps in the field. The results of the simulations suggest that this model is a very good one, and it is this that motivates our choice of mass model. This basis set allows a continuously multi-scale mass map to be reconstructed, with the angular resolution reflecting the local data quality and signal strength, but also the expected density profile cusps and slopes. We anticipate that such a basis set will be much better able to cope with the high level of noise in the data, provided that the data themselves are used to select the appropriate number of mass components used in the inference: if the weak shear data only support the inference of a small number of parameters associated with a small number of mass components then we must be able to quantify this statement, and then automatically prevent the over-fitting that can plague pixel-based methods.

While it is the NFW component parameters that are inferred given the data and the parameter priors, it is a reconstructed projected mass map that best encapsulates our state of knowledge of the cluster potential. Such maps can be constructed by tabulating the inferred (shifted and scaled) basis functions onto a grid of pixels. These maps will have, by design, highly correlated pixel values: the covariance of the pixel values should also be calculated to make the maps quantitatively useful.

The ideas introduced above, whilst not previously applied in the field of weak gravitational lensing, are not new to the inferential science community. Such "atomic" methods were suggested and developed for image reconstruction by Skilling (1998), who was motivated to move beyond pixellised models when analysing spectral and image data for the same reasons as outlined above. However, to have such a natural choice of "atom" as proposed above for galaxy clusters is something of a rare treat. However, we should remain open to the idea that the details of the atom properties is best also determined from the data – how else would we learn that the numerical simulations are realistic? In this work we demonstrate the use of NFW halos in modelling weak lensing data: alternative models to not have such well-defined prior distributions, which puts them at a natural disadvantage when comparing models. However, if a particular dataset demands a different profile atom then this can be straightforwardly inferred from the data (Kneib et al. 2003).

The methodology in this work can be rightly seen as an extension of the mass modelling of Kneib et al. (1996,and subsequent works). In this approach, one or two smooth elliptical mass components are used to model the positions and fluxes of strongly lensed images, with the parameters of the components optimised and the model refined as more multiple image systems are identified; the weak lensing data is used as a weak constraint on the strong lens mass components. Here, we adopt and justify the same modelling philosophy, but focus on the weak lensing effects of lower mass substructure at larger radii, increasing the number of free parameters, automating their estimation and pushing the interpretation of the mass components beyond that of simply stating a best-fit parameter set. Indeed, our method is much closer to the "smooth particle inference" approach put forward by Peterson et al. (2005): the differences in this case arise from the much lower signal-to noise weak lensing data (and the correspondingly fewer parameters the data can support), and the more obvious choice of basis set.

Having introduced the relevant concepts, we present in Section 2 a detailed description of the application of the atomic inference technique to weak gravitational lensing data and demonstrate its performance on simulated data in Section 4; its application to HST data is presented in (Kneib et al. 2003) (to some extent) and in (Jaunsen et al. 2005).

2 METHODOLOGY

In this section we flesh out the details of the approach introduced above.

2.1 Weak lensing background

The data considered here are the ellipticities of N background galaxies; under the assumption of intrinsically randomly oriented galaxies, whose ellipticity magnitude follows some assumed distribution, the average elliptcity provides a (noisy) estimate of the local gravitational reduced shear g (see the appendix for its definition). In practice, each of $2N$ lensed ellipticity components ϵ_j are assumed to have been drawn independently from a Gaussian distribution with mean g_j and variance $\sigma^2_{\rm intrinsic}$; here g_j is the true value of the $j^{\rm th}$ component of the (complex) reduced shear at the position of the galaxy. The likelihood function (M02) can then be written as

$$\Pr(\boldsymbol{d}|\Sigma) = \frac{1}{Z_L}\exp\left(-\frac{\chi^2}{2}\right), \tag{1}$$

where $\boldsymbol{d}$ is the vector of ellipticity (component) values, and Σ is the surface mass density used to calculate the shear fields at the background galaxy positions. χ^2 is the usual misfit statistic

$$\chi^2 = \sum_{i=1}^{N}\sum_{j=1}^{2}\frac{(\epsilon_{j,i}-g_{j,i})^2}{\sigma^2}, \tag{2}$$

and the normalisation factor is

$$Z_L = (2\pi\sigma^2)^{\frac{2N}{2}}. \tag{3}$$

The effect of errors introduced by the galaxy shape estimation procedure have been included by adding them in quadrature to the intrinsic elipticity dispersion (Hoekstra et al. 2000),

$$\sigma = \sqrt{\sigma^2_{\rm obs}+\sigma^2_{\rm intrinsic}}. \tag{4}$$

This approximation rests on the assumption that both the shape estimation error distribution is also well-approximated by Gaussians, and that the applied reduced shear is not too large. When predicting ellipticities of images close to the critical regions of a strong lensing cluster, this likelihood may be modified to relax this last assumption, a procedure suggested by Schneider et al. (2000) and implemented by Bradač et al. (2004). In practice very few galaxies are affected by this correction, and the likelihood of equation 1 is sufficiently accurate for our purposes. Any arclets lying within the critical curves of the cluster act as estimators for $1/g^*$ (*e.g.* Bradač et al. 2005): we make this correction when calculating the predicted data.

For a mass distribution Σ composed of n "atoms", the reduced shear at $\boldsymbol{\theta}$ is given by

$$g(\boldsymbol{\theta}) = \frac{\sum_1^n \gamma(\boldsymbol{\theta})}{1-\sum_1^n \kappa(\boldsymbol{\theta})}, \tag{5}$$

where γ and κ are the shear and convergence due to each mass component, *e.g.*

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_{{\rm crit},i}}. \tag{6}$$

The (lens and source redshift-dependent) critical density for the $i^{\rm th}$ galaxy is $\Sigma_{{\rm crit},i}$ – any redshift information can be included here, although non-neglible redshift errors would need to be absorbed into the likelihood function, broadening it somewhat.

2.2 The atomic (halo) mass model

The dependence on the mass distribution model parameters of the reduced shear is typically non-linear: in this general case the likelihood (equation 1) will need to be constructed from its constituent parts terms during each inference iteration. This loop over galaxy ellipticities is unavoidable, but the calculation of predicted data can be sped up by using models with analytic forms for their reduced shear fields. The NFW profile, whose choice is justified from the repeated findings of many independent N-body simulation efforts, is just such a model (Navarro et al. 1997): in the spherically symmetric case the NFW density profile is

$$\rho(r) = \frac{\rho_s}{(r/r_s)\,(1+r/r_s)^2}, \tag{7}$$

where r_s and ρ_s are the radius and density at which the logarithmic slope breaks from -1 to -3. It is useful to normalise this profile, which we treat as a two-parameter fitting function, to the mass contained within a region of overdensity 200 relative to the critical density at that redshift (Allen et al. 2003; Evrard et al. 2002):

$$\frac{M(r_{200})}{\frac{4}{3}\pi r_{200}^3} = 200\rho_{\rm crit} \tag{8}$$

$$= 4\pi\rho_s r_s^3\left[\log(1+c)-\frac{c}{1+c}\right]. \tag{9}$$

Here, $c = r_{200}/r_s$ is a measure of the concentration of the halo. Lensing properties of the NFW model have been worked out by a number of authors (Bartelmann 1996; Wright & Brainerd 2000; Meneghetti et al. 2003) and we do not reproduce their results here.

It has been found in many previous lensing analyses that it is rather important to include the ellipticity of the lens (see *e.g.* Sand et al. 2004; Kochanek et al. 2004a). There is some choice as to whether the lens potential, the deflection angle or the surface density should be of this form – since they differ by the number of times the gradient operator has been applied (0, 1 and 2 times respectively), only one of them can have concentric elliptical contours. King & Schneider (2001) chose the surface density, while Golse & Kneib (2002) opted for the slightly rectangular projected mass contours of a concentric elliptical deflection angle distribution. Citing analytic tractability, Meneghetti et al. (2003) used the lens potential $\psi(\theta')$: the major disadvantage of this approach is that for axis ratios of less than ≈ 0.7 the corresponding mass distribution becomes dumbbell-shaped. However, massive cluster potentials are likely to be close to spherical: we quantify this point below. Moreover, generally speaking (and with one eye on future joint analyses) it is perhaps better to try to characterise the shape of the potential since it is this, in equilibrium, that determines the shape of the cluster gas distribution (Buote & Canizares 1994). The breaking of axial symmetry means that the derivatives of the lens potential must be calculated explicitly in order to calculate the shear and convergence: this calculation is performed in the appendix.

N-body simulations can be used to provide prior probability distributions for the NFW profile parameters. For example, Figure 1 shows the prior on the elliptical axis ratio for a massive cluster at redshift 0.5; this pdf was derived from the fitting functions of Jing & Suto (2002), who showed

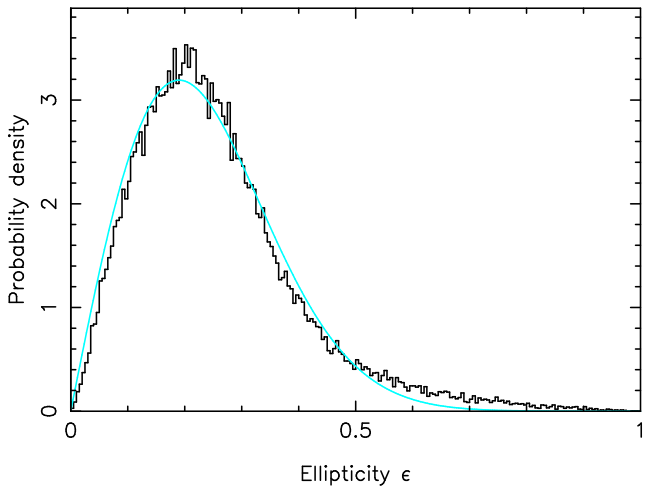

Figure 1. Lens component ellipticity prior, derived from fits to a statistical sample of N-body simulated clusters by Jing & Suto (2002). The smooth grey curve is a Rayleigh distribution of mean 0.19.

that numerically simulated cluster-scale halos can be reasonably well modelled using triaxial ellipsoidal density profiles, and tabulated the distributions of axis ratios. As indicated above, we prefer to use the more readily calculated elliptical lens potential: we use the following procedure to derive an *approximate* prior on the lens potential ellipticity parameter from the distributions tabulated by Jing & Suto, and emphasise that the estimation of this prior need only be approximate. We draw a large number of haloes from this joint pdf, and convert the axis ratios of the elliptical mass distributions in to those pertinent to the gravitational potential by applying the correction $(b/a)_\psi \approx (b/a)_\rho^{2/3}$(Kassiola & Kovner 1993). Here the subscripts ψ and ρ refer to the potential and density respectively. This equation was derived for the case of an isothermal potential and mass distribution, but we might expect it to be changed little in the case of the NFW profile. We then project the resulting isopotentials and measure the resulting projected axis ratio. This is done in an approximate way using a Gaussian form for the profile - again, we might expect the results to change little when using an NFW profile. The resulting derived prior distribution for the lens ellipticity is given in Figure 1 it is reasonably well-approximated by a Rayleigh distribution with mean 0.19, and it is this that we use as our ellipticity prior. Note that this is quite a broad prior: as such it is robust to the gross approximations and assumptions that went into its construction, and the results of the inference should depend only weakly upon it. Its main effects are to suppress unphysically high ellipticity haloes, and to favour the non-spherical haloes which are more often seen in the simulations. The position angle of the halo is assumed to follow a uniform distribution between 0 and 180 degrees: in the two-dimensional ellipticity component space the prior pdf peaks at the circularly symmetric model.

The shear due to an ellipsoidal potential is constructed by adapting the axisymmetric potential equations, altering the radius parameter in such a way as to keep the mass with in a given circular radius constant as the ellipticity changes (*e.g.* Kassiola & Kovner 1993; Meneghetti et al. 2003). The details of this process are given in the appendix. Each mass component contributes to the shear field: we assign a uniform prior distribution to the component positions within the observation region.

The priors on the NFW profile parameters may also be derived from N-body simulations. For example, Jing & Suto (2002) find the concentration parameter c to be distributed log-normally with width ≈ 0.3 about a value of 3 or so, for a massive cluster at redshift 0.5. We note that this width corresponds to something like an uncertainty of 50%. Although much larger values of the concentration have been inferred in some lensing analyses (Kneib et al. 2003; Gavazzi et al. 2003), it is not clear that the data have enough constraining power in the critical region to support this conclusion. In any case, a concentration of 20 is less than 3-sigma from the mean of the lognormal distribution, suggesting that this prior is flexible enough to cope with unusual concentrations (whilst retaining the desired feature of rejecting very low, and indeed negative, values).

Finally we consider the prior on the atom mass itself, M_{200}. The Press-Schecter formalism (Press & Schechter 1974), or one of its many numerically-corrected forms (*e.g.* Jenkins et al. 2001; Evrard et al. 2002), would serve to provide an approximate prior pdf for the halo mass were we observing a random patch of sky. The logarithmic slope of the mass function at the mass scale of galaxy clusters is close to -2, indicating the rareness of massive clusters. However, we are typically interested in pre-selected clusters, whose masses are large and typically estimable to within an order of magnitude. We suggest a compromise, and assign a Jeffreys prior pdf (logarithmic slope -1) to the halo mass, sampling uniformly in the logarithm of the mass. This has the pleasant effect of suppressing the introduction of mass into the map unless it is required by the data, but is not so severe that the halos either have masses that are heavily biased low, or are discouraged completely. A more rigorous approach would be to use the predicted halo occupation distribution to provide a joint prior on the number and mass of sub-halos, given a main cluster component. This presents some computational difficulties and is beyond the scope of this paper: the specified priors suffice to define a robust data model.

2.3 Halo parameter inference

Having defined the atomic model as the preferred representation of a weak lensing cluster, we note that it predicts data as non-linear functions of the parameters. Whilst linear methods are often favoured on the grounds of computation speed and ease of error propagation, we argue that the benefits of a fully non-linear fit outweigh these factors. We are seeking an optimal mass reconstruction, folding in as much information as we can: we do not wish to compromise this goal in favour of a computationally easier alternative. With the non-linearity comes flexibility: introducing new constraints on the mass distribution can be done in a conceptually straightforward way, adding extra terms to the posterior pdf for the model parameters. We return to this in section 3.4.

The Gaussian approximation to the posterior pdf to which we are so accustomed when using linear models may not be available in the atomic modelling. The global pos-

terior maximum is correspondingly harder to find, in a parameter space whose dimensionality can run well into double figures when many components are used. To solve these problems we employ a Markov Chain Monte Carlo sampler to explore the parameter space. This technique is now widespread in astronomical data analysis (see*e.g.* Knox et al. 2001; Lewis & Bridle 2002; Marshall et al. 2003; Dunkley et al. 2005; Bonamente et al. 2004; Peterson et al. 2005,for some examples of its use). Good introductions to the technique are given by Gilks et al. (1996) and McKay (2003): here we make the following brief comments.

Having defined a likelihood function $\Pr(\boldsymbol{d}|\eta, N)$, and sensible priors $\Pr(\eta)$ on the mass component parameters η, we note that the distribution containing all the information about the mass distribution is the posterior pdf:

$$\Pr(\eta|\boldsymbol{d}) = \frac{\Pr(\boldsymbol{d}|\eta, N)\Pr(\eta)}{\Pr(\boldsymbol{d}|N)} \tag{10}$$

Calculating the numerator of the right hand side on a fine grid throughout the parameter space would allow the normalising evidence $\Pr(\boldsymbol{d}|N)$ to be calculated, the regions of high probability to be located, and any uninteresting parameters to be numerically integrated over. In any more than a few dimensions both these operations are computationally unfeasible. It is much more convenient to work with samples drawn from the posterior distribution: both marginalisation and changing variables are trivial, the latter being done on a sample by sample basis. MCMC provides an efficient way of drawing these samples. Note the dependence of the likelihood and evidence on the number of atoms used, n (the priors on each atom's parameters having been chosen to be independent of n). The probability distribution for the atom number can be seen to be available from the data via the evidence: $\Pr(n|\boldsymbol{d}) \propto \Pr(\boldsymbol{d}|n)\Pr(n)$. With the (generous) assignment of a uniform prior on the atom number, the evidence gives the (discrete) pdf for n directly. In the same way the evidence may also be used to quantify he relative probabilities of two or more competing atom models, a process carried out in (Kneib et al. 2003). The evidence lies at the heart of all Bayesian model selection and hypothesis testing(see *e.g.* MacKay 2003,for an excellent introduction), and its use is growing in astronomy (*e.g.* Jaffe 1996; Knox et al. 1998; Hobson et al. 2002; Marshall et al. 2003; Mukherjee et al. 2005).

We use the freely available software package `bayesys3`, written by John Skilling. This general purpose code, used in previous work on this subject (Marshall et al. 2003), is known to cope well with the types of likelihood surface presented by weak lensing datasets: the evidence values are found to be accurate and their calculation readily repeated. The evidence is calculated by thermodynamic integration during the burn-in period (Ó Ruanaidh & Fitzgerald 1996).

The only disadvantage to using MCMC rather than an optimisation followed by a Gaussian approximation to the posterior is that it can be slow: typically, analysing a catalogue of some few thousand galaxies using a model consisting of three components on a 3GHz processor can be expected to take several hours. This should be compared with the time taken by alternative methods to find not only the optimised model parameters, but also estimates of their uncertainties.

2.4 Probabilistic mass mapping

While the parameters of individual halos may well be of interest (*e.g.* Kneib et al. 2003), the information on the mass distribution can be displayed in a more visually helpful way, in the form of a mass image. Each MCMC sample corresponds to a set af halos that provide an acceptable fit to the data: the projected mass distribution of these halos can be mapped on to a pixellised grid. The probability distribution of the surface mass density in any given pixel can be built up by calculating its value for each sample, and forming a histogram. This is the pdf marginalised over all halo parameters, so the width of this distribution represents the maximum uncertainty of that pixel value, given the assumptions. To make a map, the individual pixel probability distributions have to be reduced to one number: we use the arithmetic mean for its ease of calculation (Skilling 1998), but note that an alternative central value may be more appropriate if the pixel value pdf is highly skewed.

The resulting reconstructed maps inevitably retain some of the appearance of the halos from which they are composed: however, averaging over the posterior pdf does bring out some extra information not present in any individual sample map. These reconstructions can be thought of as being highly regularised – by design, given the noisy data – using a multi-scale kernel whose shape has been chosen to be appropriate for dark matter in cluster halos. In the next section we show some examples of these atomic maps, and how well they describe the lensing data.

3 DEMONSTRATION ON SIMULATED DATA

In order to demonstrate the methodology introduced above, we generate mock weak lensing data for a typical, moderately massive ($M_{\rm virial} \approx 6 \times 10^{14} M_\odot$), unrelaxed cluster at redshift 0.55. This cluster is an N-body simulation, from the sample in Eke, Navarro & Frenk (1998). We place background galaxies at random positions on a single source plane at redshift 1.2, such that the cluster has a small critical region (radius 5 arcsec) with a number density of 80 per square arcminute over a square field 3 arcminutes on a side. This was intended to represent a standard weak lensing dataset from the ACS camera on HST, and resulted in a catalogue of 741 galaxy shapes. An intrinsic ellipticity distribution of width 0.25, and shape estimation error of 0.2, were assumed; the applied reduced shear was calculated by convolving the mass distribution with the lensing kernel and scaling by the critical density as in M02. We note that this process used a projected density map with pixel scale ≈ 2 arcsec, fine enough to retain any cuspy features in the mass distribution. When assessing the statistical performance of the atomic inference algorithm, multiple noise realisations were used. Each realisation corresponds to a separate set of background galaxy positions and intrinsic ellipticities, and a different shape estimation noise term. The true projected mass distribution, scaled by the critical density for this lens and source redshift, is shown in Figure 2, along with one realisation of the simulated weak lensing data.

A single source plane was employed for simplicity, and to separate out the performance of the atomic modelling from systematic effects due to unknown background galaxy

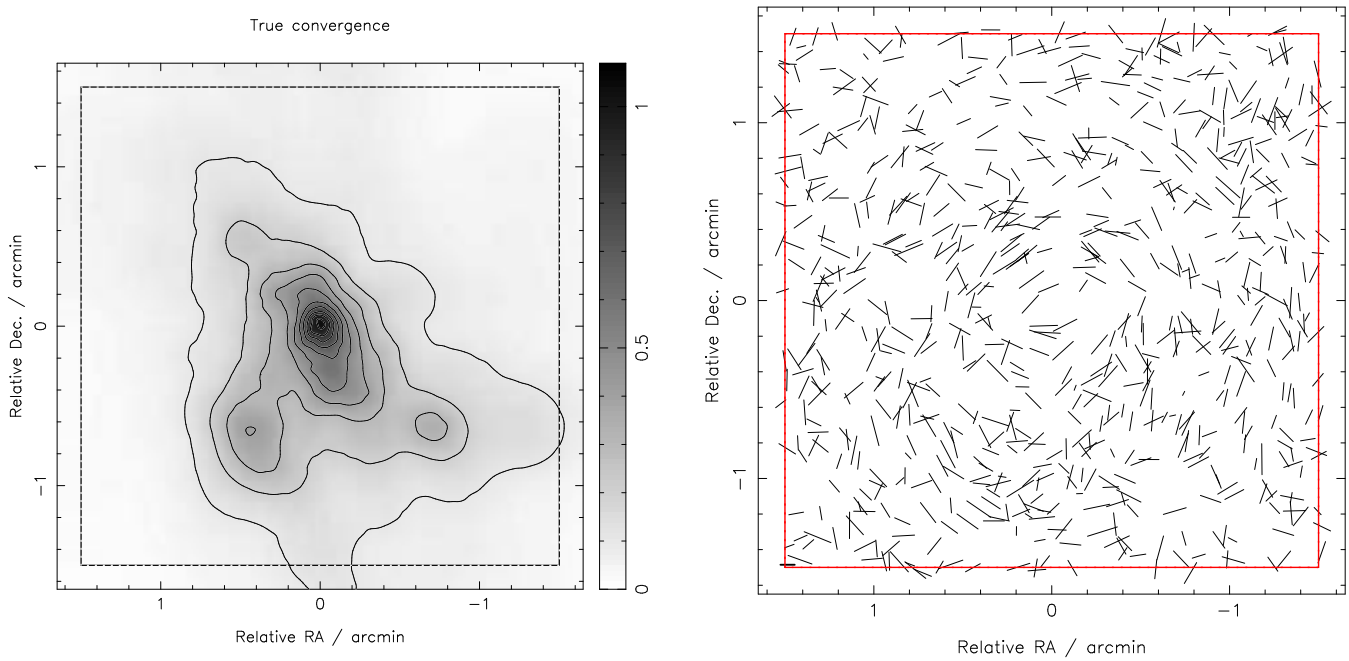

Figure 2. True convergence distribution (left) and reduced shear field (right) for the $z = 0.55$ N-body simulated cluster CL09, assuming a source plane redshift of 1.2. The 3-arcminute square observing region is shown by the dashed box. One main clump, and two minor clumps, are visible. A typical realisation of the sampling of the reduced shear field by the galaxy shapes is shown.

redshifts. We do note, however, that including measured redshifts for each source is trivial in this algorithm, since we predict the reduced shear on each object.

3.1 Estimating the number of haloes

We first investigate the number of halos appropriate for modelling this dataset. This was done by sampling the posterior distributions of the parameters of an n-halo model, where n was allowed to increase from 1 to 5. For each inference, the evidence was calculated, and is shown in Figure 3. These evidence values are readily reproducible, as indicated by the error bars on the plot: these show the standard deviations of the mean log evidence over 5 runs of the sampler, and are routinely less than one unit. The error bars reflect the scatter between different noise realisations, of which ten are overlaid in the figure. This plot gives some confidence in the ability of the sampler to simulate the posterior pdf; it also quantifies the quality of the data available in observations such as that simulated.

The evidence for the n-halo model is $\Pr(\boldsymbol{d}|n)$: the ratio of this to the evidence for zero atoms (*i.e.*the null model, with zero predicted surface mass density) gives a measure of the significance of the detection (Hobson & McLachlan 2003). Were Figure 3 to be plotted over a wider range in the ordinate, probability (evidence) ratios relative to the null model of e^{40-60} would have been visible, indicating a resounding detection of the cluster in the shear data. The differences between evidence values for n of 1, 2, 3, 4 and 5 components are much smaller, typically reaching between two and four units in the logarithm between the peak evidence and the $n = 5$ value. These differences correspond to probability ratios of approximately $10 - 50$: the data are found to be typically an order of magnitude more likely to

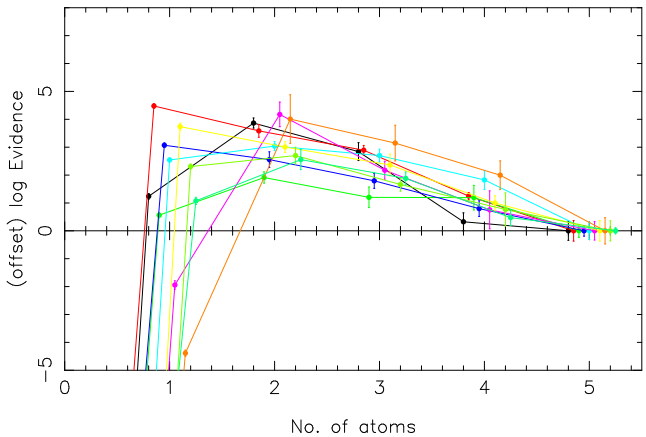

Figure 3. (Renormalised) log evidence as a function of halo number for the simulated cluster CL09. The points with error bars show the evidence estimates at each atom number: the ten lines correspond to ten different noise realisations of the simulation. A broad peak around atom number 1-3 is typically seen.

have come from a two-component model than a 5-component model (with the exact favoured value of n being dependent on the noise realisation). This is an important result: the simulated data were designed to be representative of that being analysed in contemporary work; what we are seeing here is a quantification of the amount of information available to us from that data. With the well-appointed halo model used, only a handful of parameters are both required and supported by the data. This is in agreement with the findings of M02: in the next section we show how the inferred mass distributions differ from the maximum entropy maps.

3.2 Mass mapping

The left-hand column of Figure 4 shows the ensemble-averaged halo-model mass reconstructions, for two different noise realisations. For comparsion, the maximum-entropy maps are shown along side, at two different resolution scales. These plots serve to illustrate some general points about the two techniques. The lensent2 algorithm, and all other methods that use a single resolution scale when smoothing the data or in the reconstruction process itself, does not cope well with the rangel of scales of the mass structure in this cluster. A sharp, cuspy peak, surrounded by an extended irregular mass distributions in the outer regions clearly requires at least two scales (approximted here by the 15 and 40-arcsec lensent2 resolution kernels): these scales are provided naturally by our chosen components in the atomic maps. The atomic reconstruction is remarkably robust between noise realisations, whereas the more flexible pixel-based method maps contain tranisent structure as the noise is fitted. This problem was alleviated in lensent2 by increasing the resolution scale until, at maximum evidence, only believable features remained; the same may be said about the structure in the atomic maps, except that the small-scale, high signal-to-noise structure is retained (*e.g.*at the centre of the cluster). Finally we note the long standing problem of inferring super-critical density from weak shear data, outlined in some detail in the work of Schneider & Seitz (1995; 1995). With no additional information it is impossible to infer uniquely the presence of convergence greater than unity: when constructing the mass distribution from naturally cuspy components the convergence can be effectively interpolated upwards in a seamless way.

The observations of the previous paragraph can be put on a more quantitative footing by plotting the correlation between the inputs and outputs of the algorithms. This is shown in Figure 5. We use the correlation function $\xi_+(\theta)$, where θ is the angular separation between galaxy pairs; this is given by (*e.g.* Kochanek et al. 2004b):

$$\xi_+(\theta) = \langle g_{\rm t}^{\rm A} g_{\rm t}^{\rm B}\rangle + \langle g_{\rm x}^{\rm A} g_{\rm x}^{\rm B}\rangle, \tag{11}$$

where $g_{\rm t}^{\rm A}$ ($g_{\rm x}^{\rm A}$) is the tangential (radial) component of the reduced shear estimator at galaxy position of galaxy A, relative to galaxy position B (and vice versa), and A and B are separated by θ. This function conveniently quantifies the alignment of pairs of galaxy shapes. Since we are interested in the difference between the reduced shear predicted by the reconstructions (g), and either the true reduced shear ($\hat{g}$) or measured ellipticities (ϵ), we construct the difference function

$$\Delta\xi_+(\theta) = \langle (g_{\rm t}^{\rm A} - \hat{g}_{\rm t}^{\rm A})(g_{\rm t}^{\rm B} - \hat{g}_{\rm t}^{\rm B})\rangle + \langle (g_{\rm x}^{\rm A} - \hat{g}_{\rm x}^{\rm A})(g_{\rm x}^{\rm B} - \hat{g}_{\rm x}^{\rm B})\rangle \tag{12}$$

$$= \xi_+^{gg}(\theta) + \xi_+^{\hat{g}\hat{g}}(\theta) - 2\xi_+^{g\hat{g}}(\theta). \tag{13}$$

For a perfect match on all scales, this function whould be zero. All correlation functions decrease with increasing pair separation, as the lensing signal diminishes in strength. The upper panel of Figure 5 shows that the residuals in the 3-atom reconstruction, and the evidence-preferred 40-arcsec lensent2 reconstruction, are consistent with noise; the high resolution lensent2 reconstruction shows a small positive $\Delta\xi_+(\theta)$ at scales of 5-40 arcsec indicative of an imperfect reconstruction. This shortcoming is seen more clearly in the lower panel; in the figure the low resolution lensent2 map and the 3-atom reconstruction are seen to do roughly equally well in recovering the true mass distribution, with the atomic map doing slightly better on the smallest scales. This agrees with the maps of Figure 4, where the cuspy cluster centre is not reproduced with the smooth maximum-entropy map.

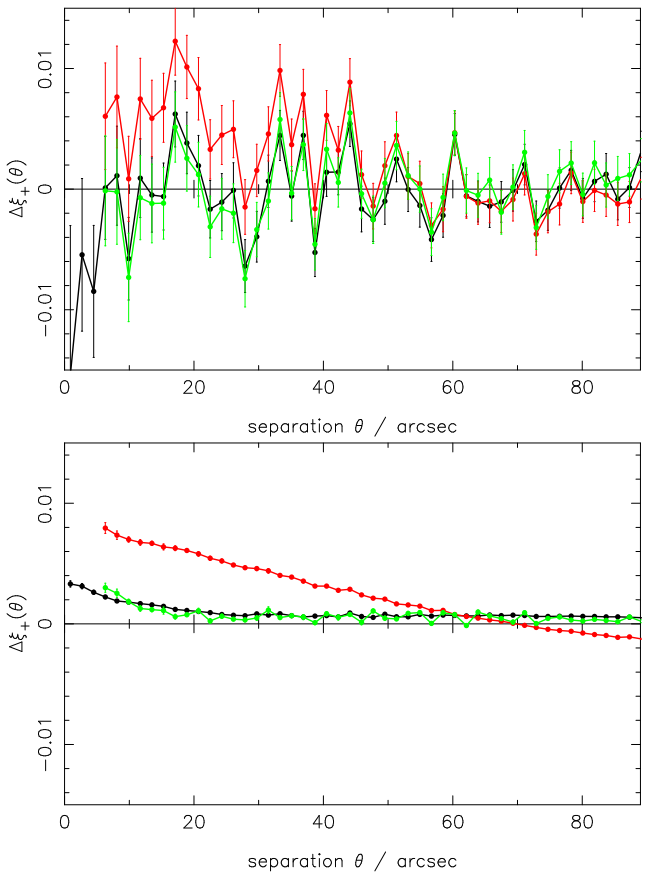

Figure 5. Quantifying the mass map reconstruction accuracy via the correlation function differences. Top: correlation function ξ_+ of the difference between the ensemble-average predicted reduced shear, and the data. The curves shown are for the 3-halo atomic reconstruction (black), 15-arcsec resolution lensent2 map (red), and the 40-arcsec lensent2 map (green). Bottom: the same exercise, with the same legend, but now comparing the predicted reduced shear with the true input values.

3.3 Component properties

The ensemble average mass map is one way of representing the information in the joint posterior pdf; marginal distributions for other parameters of interest are also readily available. For example, Figure 6 shows the position, and mass profile parameters associated with the secondary subclump visible in the projected mass map. MCMC samples corresponding to a circular region centred on the map peak were excised and histogrammed, to plot the distributions $\Pr(x, y|\boldsymbol{d}, H)$ and $\Pr(M_{200}, c|\boldsymbol{d}, H)$, where H is the assumption that a mass feature of interest lies within this aperture. The widths of these distributions provide estimates of the uncertainties on the parameters, obtained without the need for bootstrap re-sampling or simulation of mock data – the MCMC process performed exactly this calculation during

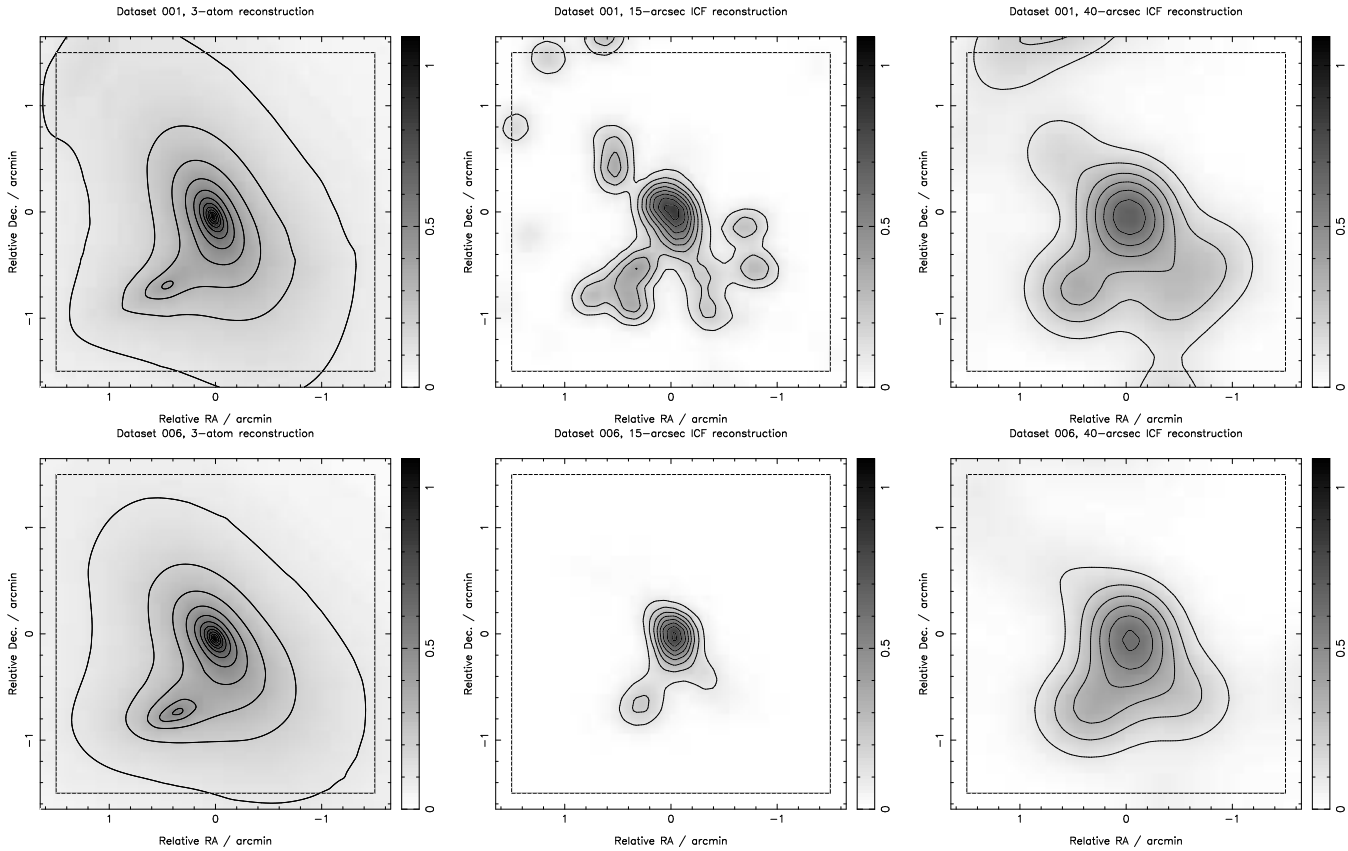

Figure 4. Two representative mass reconstructions, one noise realisation per row, for the $z = 0.55$ N-body simulated cluster CL09. Left column: ensemble-average atomic inference mass distributions. Centre and right columns: 15 and 40-arcsec resolution maximum-entropy maps These panels may be compared directly with the left-hand panel of figure 2. The 3-arcminute square observing region is shown by the dashed box.

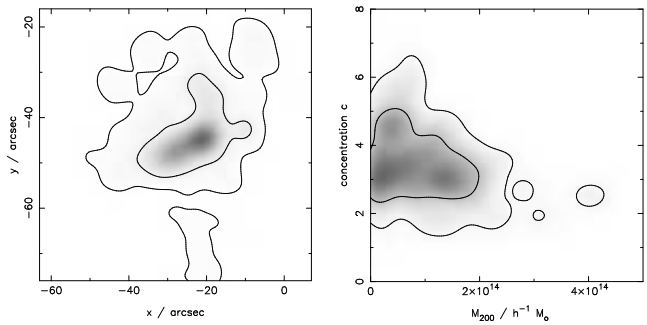

Figure 6. Inferences about the south-eastern substructure seen in the reconstruction of Figure 4. Top: $\Pr(x, y|\boldsymbol{d})$; bottom: $\Pr(M_{200}, c|\boldsymbol{d})$. In the right panel the prior on the concentration can be seen, allowing a reasonable estimate of the mass of the marginally-detected subclump.

the inference. This secondary feature is something of a transient one: in some noise realisations the signal is too broken up to be detected.

3.4 Including strong lensing

The non-linear nature of the inference process described in the previous sections makes it straightforward to include strong lensing constraints as priors. For example, the results of an independent strong lens modelling may typically yield estimates ($e.g. \sigma_\kappa$), with error bars, of the convergence (κ) and shear at a particular point. These uncertainties, when translated into Gaussian (or the appropriate error distribution) pdfs centred on the point estimate ($e.g. \kappa_0$), can be used to weight the MCMC trial parameter sets: the weak lensing likelihood is simply multiplied by the value of a pdf such as $\Pr(\kappa|\kappa_0, \sigma_\kappa)$.

Note that we make no distinction between cluster-scale and galaxy-scale strong lensing: the results of the analysis of both types of multiple-image systems can be presented as the aforementioned pdfs. However, multiple image systems in clusters contain information on the details of the mass distribution, and one would have to be careful that the distributions for κ propagated this information in full. A more rigorous approach would be closer to that of Kneib et al. (1996), whereby the properties of the multiple images are fitted directly: such an analysis is beyond the scope of this work.

In the simulation described above, a galaxy-scale strong lensing constraint was placed on the reconstruction, corresponding to an unbiased strong lens estimate of the "external" convergence and shear at various points near the south-western (undetected) sub-clump visible in Figure 2. With realistic error estimates ($\sigma_\kappa = \sigma_{\gamma_i} = 0.02$, the effect on the reconstruction is small: the third mass feature is hinted at in some noise realisations' reconstructions, and the evidence peak shifts towards slightly large numbers of atoms, but the

basic result of only a few mass components being required is unchanged. With lower signal-to-noise data we might expect this kind of constraint to have more of an effect. Likewise, the multiple constraints arising form a cluster lensing analysis would have a bigger effect on the reconstruction, as shown in Kneib et al. (2003).

The principal result of this work is the low evidence for large numbers of mass components in weak lensing maps. Having shown that the atomic inference maps provide good quality reproductions of the underlying mass distribution, we can now investigate the information available in a weak lensing dataset for predicting the convergence and shear at specific points in the field. Were such a prediction possible with high precision, measuring Hubbles' constant from time delays in strong lenses in cluster environments (*e.g.* Koopmans et al. 2003) would become a competitive alternative to other methods. Plotted in the left panel of Figure 7 is the difference between the top reconstruction of Figure 4 and the true mass distribution; for comparison, the central panel shows the width of the pixel pdf, as estimated by the standard deviation of the samples. This "error" map is informative: even in the regions where the shear signal is strong, the uncertainty on the predicted convergence is high. However, using this uncertainty map to rescale the residuals between reconstruction and truth we see that the differences are fairly low significance: only in the region of the undetected sub-clump are the pixel values more than 3 sigma (where sigma is the uncertainty mapped in the central panel) from the truth.

Figure 7 illustrates the small-scale mass structure in the cluster core, undetectable by weak lensing, that makes pointwise convergence prediction at the level of 0.02 or better very difficult. Clearly more information is needed: the work of Natarajan & Springel (2004) indicates that including the mass associated with galaxies via small atoms placed at the cluster member positions in order to approximate this substructure and obtain meaningful convergence estimates. Investigating the accuracy of this approach is beyond the scope of this paper, but would be an important step in understanding the strong lens time delays.

4 DISCUSSION

Weak lensing data is very noisy: in such situations, the prior pdf on the parameters of any model we choose to fit is likely to be comparable in importance as the likelihood function. Admitting this, we have translated the information on mass distributions in clusters of galaxies available from numerical simulations and translated this into a sensible set of priors. The atomic inference maps then include that information in a natural way. We point out explicitly that these mass maps are biased, in the sense that no matter how good the data is the predicted mass distribution still has to be constructed (within this particular model) from NFW-shaped halos, and in any individual cluster observed at a high signal-to-noise ratio one can imagine this model breaking down. Indeed, careful inspection of the maps in Figure 7 shows that the reconstructed surface density is systematically higher than the true density in the outer parts of the field of view, by 0.05 in convergence or so. This is despite the ellipticity data being fitted to a fully satisfactory level (Figure 5), and is due to the mass sheet degeneracy (Falco et al. 1985). This effect when working with parameterised profiles was pointed out by Schneider et al. (2000) and further investigated by Bradač et al. (2004), and was left as a limitation on the measurement of cluster density profiles. The converse of this is that if one has additional information on the profile, then one can break the mass sheet degeneracy: more precisely, in the present case the mass sheet degeneracy effect is incorporated it into the model-dependent error bars available from the analysis of the MCMC samples.

The model-dependence of the inferences advocated here has been billed as a sensible use of the prior information available from N-body simulations. However, one should remember that the universal profiles are average profiles of halos: in any given cluster it may be that some alternative profile is more appropriate, and that the prior distribution for the concentration may not afford the atom profiles sufficient freedom to fit the data well. In this case, the profile can be inferred from the data via the evidence as was shown in (Kneib et al. 2003). When using any other profile, the priors on the halo parameters will not be as readily derived, and indeed may be preferred to be kept uninformative. In this case the Occam's razor factor inherent to the evidence will act to favour the basic (and better-constrained) NFW atom set. A consequent increase in evidence when using the alternative profile will then be rather a robust conclusion, the analyst's natural tendency towards the expected forms being taken care of already.

5 CONCLUSIONS

We have presented a new method for reconstructing the mass distribution of clusters of galaxies from weak gravitational lensing data, based on the atomic inference procedure suggested by Skilling (1998). By investigating its performance on simulated HST ACS data, we draw the following conclusions:

- Perhaps as expected, the number of mass components supported by the data (and selected via the evidence) is typically quite small. This is a consequence of the domination of clusters by a single deep potential with a few satellite haloes, combined with the paucity of information on the weak shear data.
- The atomic inference (posterior average) mass maps are cleaner and more robust to the noise realisation, than those made with pixel-based methods: the natural basis set of elliptical NFW profile halos provides something like optimal regularisation, suppressing spurious peaks and providing an accurate reconstruction. This accuracy is evident in the high-strength correlation between input map and reconstruction, that extends (by virtue of the additional information input to the model) to smaller angular scales than previous techniques. The maps are biased towards the results of numerical simulations, incorporating our expectations of what clusters should look like. As a result, the mass sheet degeneracy is broken, and absorbed into the uncertainties associated with the maps.
- The necessarily non-linear method, whilst slow to execute, has the advatage of easily accommodating additional physical constraints of an arbitrary functional form: the independent estimate of convergence and shear at a point im-

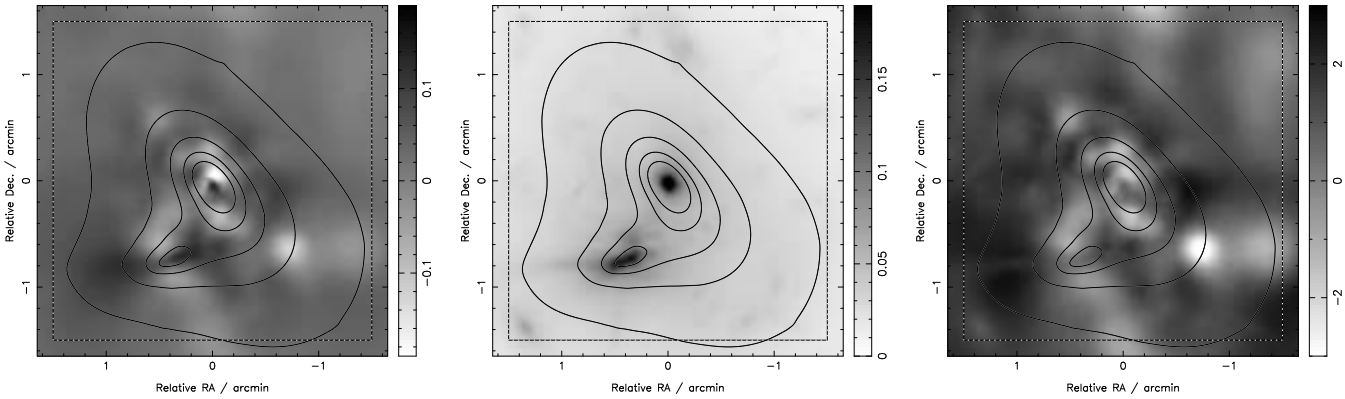

Figure 7. Left: difference map between the 3-halo reconstruction of Figure 4 and the true surface density seen in Figure 2. Centre: the atomic uncertainty map, as estimated by the standard deviations of the individual pixel pdfs. Right: the difference map, divided by the error map. In all panels the reconstructed convergence contours are overplotted to guide the eye.

proves the accuracy of the reconstruction, an effect that is especially strong with low signal-to-noise data.

This last point opens the door for a more comprehensive weak plus strong lensing reconstruction method. There are two challenges here: firstly, solving for image positions, or better, predicting entire lensed images, is very time consuming: even with the small numbers of components inferred for typical clusters the cpu time borders on the edge of the analyst's patience. Secondly, most of the difficulty in strong lens modelling is associated with identifying and confirming which objects form multiple-image systems. At present this appears to be still best achieved by human skill and judgement, although with Moore's law in mind it may be worthwhile to invest in automated schemes for analysing large numbers of strong lensing clusters from future wide field surveys. The methodology outlined here is conceptually very clean, and consequently provides a framework within which additional constraints on the cluster mass distribution can be incorporated.

The (standard fortran77 and c) code used in this and the cited work is available on request from the author.

ACKNOWLEDGMENTS

We thank Mike Hobson and Steve Gull for useful discussions on atomic inference methods, and John Skilling for moreover providing his BayeSys3 code (freely available from `http://www.inference.phy.cam.ac.uk/bayesys/`). We also thank Jean-Paul Kneib, Roger Blandford and Marusa Bradac for their comments and advice, and Patrick Hudelot for helping test the McAdam code. This work was supported in part by the U.S. Department of Energy under contract number DE-AC02-76SF00515.

REFERENCES

Allen S. W., Schmidt R. W., Fabian A. C., 2002, MNRAS, 334, L11
Allen S. W., Schmidt R. W., Fabian A. C., Ebeling H., 2003, MNRAS, 342, 287
Bartelmann M., 1996, A&A, 313, 697
Bonamente M., Joy M. K., Carlstrom J. E., Reese E. D., LaRoque S. J., 2004, ApJ, 614, 56
Bradač M., Erben T., Schneider P., Hildebra ndt H., Lombardi M., Schirmer M., Miralles J.-M., Clowe D., Schindler S., 2005, A&A, 437, 49
Bradač M., Lombardi M., Schneider P., 2004, A&A, 424, 13
Bradač M., Schneider P., Lombardi M., Erben T., 2005, A&A, 437, 39
Broadhurst T., Takada M., Umetsu K., Kong X., Arimoto N., Chiba M., Futamase T., 2005, ApJL, 619, L143
Buote D. A., Canizares C. R., 1994, ApJ, 427, 86
Clowe D., Gonzalez A., Markevitch M., 2004, ApJ, 604, 596
Clowe D., Schneider P., 2002, A&A, 395, 385
Czoske O., Moore B., Kneib J.-P., Soucail G., 2002, A&A, 386, 31
Dunkley J., Bucher M., Ferreira P. G., Moodley K., Skordis C., 2005, MNRAS, 356, 925
Eke V., Navarro J., Frenk C., 1998, ApJ, 503, 569
Evrard A. E., MacFarland T. J., Couchman H. M. P., Colberg J. M., Yoshida N., White S. D. M., Jenkins A., Frenk C. S., Pearce F. R., Peacock J. A., Thomas P. A., 2002, ApJ, 573, 7
Falco E., Gorenstein M., Shapiro I., 1985, ApJ, 289, L1
Gavazzi R., Fort B., Mellier Y., Pelló R., Dantel-Fort M., 2003, A&A, 403, 11
Gilks W. R., Richardson S., Spiegelhalter D. J., 1996, Markov-Chain Monte-Carlo In Practice. Cambridge: Chapman and Hall
Golse G., Kneib J., 2002, A&A, 390, 821
Hobson M. P., Bridle S. L., Lahav O., 2002, MNRAS, 335, 377
Hobson M. P., McLachlan C., 2003, MNRAS, 338, 765
Hoekstra H., Franx M., Kuijken K., 2000, ApJ, 532, 88
Jaffe A., 1996, ApJ, 471, 24
Jaunsen A., Marshall P. J., Kneib J., Hjorth J., 2005, in preparation
Jee M. J., White R. L., Benítez N., Ford H. C., Blakeslee J. P., Rosati P., Demarco R., Illingworth G. D., 2005, ApJ, 618, 46
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

Jing Y. P., Suto Y., 2002, ApJ, 574, 538
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kassiola A., Kovner I., 1993, ApJ, 417, 450
King L. J., Schneider P., 2001, A&A, 369, 1
Kneib J., Hudelot P., Ellis R. S., Treu T., Smith G. P., Marshall P., Czoske O., Smail I., Natarajan P., 2003, ApJ, 598, 804
Kneib J.-P., Ellis R., Smail I., Couch W., Sharples R., 1996, ApJ, 471, 643
Knox L., Bond J. R., Jaffe A. H., Segal M., Charbonneau D., 1998, PRD, 58, 083004
Knox L., Christensen N., Skordis C., 2001, ApJL, 563, L95
Kochanek C. S., Schneider P., Wambsganss J., 2004a, in Meylan G., Jetzer P., North P., eds, Gravitational Lensing: Strong, Weak & Micro Lecture Notes of the 33rd Saas-Fee Advanced Course, Strong Gravitational Lensing. Springer-Verlag: Berlin
Kochanek C. S., Schneider P., Wambsganss J., 2004b, in Meylan G., Jetzer P., North P., eds, Gravitational Lensing: Strong, Weak & Micro Lecture Notes of the 33rd Saas-Fee Advanced Course, Weak Gravitational Lensing. Springer-Verlag: Berlin
Koopmans L. V. E., Treu T., Fassnacht C. D., Blandford R. D., Surpi G., 2003, ApJ, 599, 70
Lewis A., Bridle S., 2002, Phys.Rev. D, 66, 103511
Lombardi M., Rosati P., Blakeslee J. P., Ettori S., Demarco R., Ford H. C., Illingworth G. D., Clampin M., Hartig G. F., Benítez N., Broadhurst T. J., Franx M., Jee M. J., Postman M., White R. L., 2005, ApJ, 623, 42
MacKay D., 2003, Information Theory, Inference and Learning Algorithms. Cambridge: CUP
McKay D. J. C., 2003, Information Theory, Inference, and Learning Algorithms. Cambridge: CUP
Marshall P. J., Hobson M. P., Gull S. F., Bridle S. L., 2002, MNRAS, 335, 1037
Marshall P. J., Hobson M. P., Slosar A., 2003, MNRAS, 346, 489
Meneghetti M., Bartelmann M., Moscardini L., 2003, MNRAS, 340, 105
Mukherjee P., Parkinson D., Liddle A. R., 2005, ArXiv Astrophysics e-prints
Natarajan P., Springel V., 2004, ApJL, 617, L13
Navarro J., Frenk C., White S., 1997, ApJ, 490, 493
Ó Ruanaidh J., Fitzgerald W., 1996, Numerical Bayesian Methods Applied to Signal Processing. New York: Springer-Verlag
Peterson J. R., Marshall P. J., Andersson K., 2005, ApJ submitted, astro-ph/0507613
Press W., Schechter P., 1974, ApJ, 187, 425
Sand D. J., Treu T., Smith G. P., Ellis R. S., 2004, ApJ, 604, 88
Schneider P., King L., Erben T., 2000, A&A, 353, 41
Schneider P., Seitz C., 1995, A&A, 294, 411
Seitz C., Schneider P., 1995, A&A, 297, 287
Skilling J., 1998, in Erickson G., Rychert J. T., Smith C. R., eds, , Maximum Entropy and Bayesian Methods. Dordrecht: Kluwer, p. 14
Tyson J., Valdes F., Wenk R., 1990, ApJ, 349, L1
Wright C. O., Brainerd T. G., 2000, ApJ, 534, 34

APPENDIX

Starting from the lensing potential ψ, which is related to the Newtonian gravitational potential Φ by

$$\psi(\theta) = \frac{1}{2\pi G\Sigma_{\rm crit} D_\theta^2}\int \Phi(\theta D_\theta, z)dz, \tag{14}$$

the convergence and shear are given by

$$\kappa = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial\theta_1^2} + \frac{\partial^2\psi}{\partial\theta_2^2}\right), \tag{15}$$

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial\theta_1^2} - \frac{\partial^2\psi}{\partial\theta_2^2}\right), \tag{16}$$

$$\gamma_2 = \frac{\partial^2\psi}{\partial\theta_1\partial\theta_2}. \tag{17}$$

The "observable" reduced shear g is given by

$$g(\theta) = \frac{\gamma(\theta)}{1-\kappa(\theta)}, \tag{18}$$

where the complex quantity $\gamma = \gamma_1 + i\gamma_2$.

For an axisymmetric lens potential $\psi(\theta)$ one can write

$$|\gamma|(\theta) = \frac{1}{2}\left(\frac{1}{\theta}\frac{d\psi}{d\theta} - \frac{d^2\psi}{d\theta^{\,2}}\right), \tag{19}$$

with the shear components given by

$$\gamma_1 = \left(\frac{\theta_2^2-\theta_1^2}{2\theta^2}\right)|\gamma|, \tag{20}$$

$$\gamma_2 = -\frac{2\theta_1\theta_2}{\theta^2}|\gamma|, \tag{21}$$

Given analytic forms for ψ and its gradients, the reduced shear at any point$\boldsymbol{\theta} = (\theta_1, \theta_2)$ on the sky can be calculated.

Ellipticity is incorporated into the mass distribution by applying the coordinate transform

$$\begin{pmatrix}\theta_1'\\ \theta_2'\end{pmatrix} = \begin{pmatrix}\sqrt{f} & 0\\ 0 & 1/\sqrt{f}\end{pmatrix}\begin{pmatrix}\cos\phi & \sin\phi\\ -\sin\phi & \cos\phi\end{pmatrix}\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}, \tag{22}$$

which, in vector notation, reads

$$\boldsymbol{\theta'} = \mathsf{SR}\boldsymbol{\theta}. \tag{23}$$

Here, f is the ratio of semi-minor to semi-major axes, b/a, such that lines of constant θ' are ellipses enclosing area πab aligned at angle ϕ to the θ_1-axis. The elliptical radius θ' is then calculated by

$$\theta'^{\,2} = \boldsymbol{\theta}^{\rm T}\mathsf{R}^{\rm T}\mathsf{S}^{\rm T}\mathsf{SR}\boldsymbol{\theta} \tag{24}$$

$$= \theta_i\mathsf{M}_{ij}\theta_j, \tag{25}$$

where suffix notation for the components of the tensors has been adopted, and the matrix product $\mathsf{R}^{\rm T}\mathsf{S}^{\rm T}\mathsf{SR}$ has been renamed M. Functions of θ' only have contours which are concentric ellipses.

With the lens potential a function of θ' only, and with primes denoting differentiation with respect to this variable, one finds

$$\frac{\partial\psi}{\partial\theta_k} = \frac{\partial\theta'}{\partial\theta_k}\psi' \tag{26}$$

$$\frac{\partial^2\psi}{\partial\theta_i\partial\theta_j} = \frac{\partial^2\theta'}{\partial\theta_i\partial\theta_j}\psi' + \frac{\partial\theta'}{\partial\theta_i}\frac{\partial\theta'}{\partial\theta_j}\psi'' \tag{27}$$

Differentiating equation (25) once gives

$$\frac{\partial\theta'}{\partial\theta_k} = \frac{1}{2\theta'}\theta_k(\mathsf{M}_{kj} + \mathsf{M}_{jk}) \tag{28}$$

$$= \frac{1}{\theta'}\mathsf{Q}_{jk}\theta_k, \tag{29}$$

where Q is the symmetric part of M. Differentiating again gives

$$\frac{\partial^2 \theta'}{\partial \theta_i \partial \theta_j} = \frac{1}{\theta'} \left[\mathsf{Q}_{ij} - \frac{\mathsf{Q}_{ik}\theta_k \mathsf{Q}_{jl}\theta_l}{\theta'^2} \right] \tag{30}$$

Equation (27) can then be used to compute the quantities needed in equations (15–17).

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