**Astronomy
&
Astrophysics**

Strong and weak lensing united

I. The combined strong and weak lensing cluster mass reconstruction method[*]

M. Bradač[1,2,3], P. Schneider[1], M. Lombardi[1,4,5], and T. Erben[1]

[1] Institut für Astrophysik und Extraterrestrische Forschung, Auf dem Hügel 71, 53121 Bonn, Germany
    e-mail: marusa@astro.uni-bonn.de
[2] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany
[3] KIPAC, Stanford University, 2575 Sand Hill Road, Menlo Park, CA 94025, USA
[4] European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
[5] Università degli Studi di Milano, v. Celoria 16, 20133 Milano, Italy

Abstract. Weak gravitational lensing is considered to be one of the most powerful tools to study the mass and the mass distribution of galaxy clusters. However, the mass-sheet degeneracy transformation has limited its success. We present a novel method for a cluster mass reconstruction which combines weak and strong lensing information on common scales and can, as a consequence, break the mass-sheet degeneracy. We extend the weak lensing formalism to the inner parts of the cluster and combine it with the constraints from multiple image systems. We demonstrate the feasibility of the method with simulations, finding an excellent agreement between the input and reconstructed mass also on scales within and beyond the Einstein radius. Using a single multiple image system and photometric redshift information of the background sources used for weak and strong lensing analysis, we find that we are effectively able to break the mass-sheet degeneracy, therefore removing one of the main limitations on cluster mass estimates. We conclude that with high resolution (e.g. HST) imaging data the method can more accurately reconstruct cluster masses and their profiles than currently existing lensing techniques.

Key words. cosmology: dark matter – galaxies: clusters: general – gravitational lensing

1. Introduction

Clusters of galaxies have long been recognised as excellent laboratories for many cosmological applications. An especially important diagnostic is their number density as a function of mass and redshift. This can only be measured if reliable mass estimates of the observed clusters can be obtained. In addition, in the framework of the ΛCDM cosmological model, the dark matter distribution in clusters likely follows the NFW profile (Navarro et al. 1997).

Weak gravitational lensing is one of the most powerful tools currently available for studying the mass distribution of clusters of galaxies. The first weak lensing detection in clusters has been made by Tyson et al. (1990). However, it was only after the pioneering work by Kaiser & Squires (1993) that the field began to flourish, and since then many cluster mass reconstructions have been carried out (see e.g. Clowe & Schneider 2001, 2002; Gavazzi et al. 2004; Lombardi et al. 2005). A disagreement occurring in some cases between the cluster mass estimated from the weak/strong lensing measurements with X-rays is still not well understood, although several scenarios have

been proposed to resolve this issue (see e.g. Allen 1998; Ettori & Lombardi 2003).

In the absence of the redshift information, the main limitation for a precise weak lensing mass estimate is the mass-sheet degeneracy (Schneider & Seitz 1995). If the redshifts of background sources and/or lens are not known, the transformation of the surface mass-density $\kappa \rightarrow \kappa' = \lambda\kappa + (1 - \lambda)$, where $\lambda \neq 0$ is an arbitrary constant, leaves the expectation value of measured image ellipticities unchanged. In Bradač et al. (2004a) we show that this degeneracy can be lifted using information on *individual* source redshifts, however only if the weak lensing reconstruction is extended to the critical parts of the cluster. Strong lensing is affected by this transformation as well. Namely, the mass-sheet degeneracy does not change the image positions (since the source position is not an observable) and flux ratios and therefore can not be broken if a single redshift multiple-image system is used. The mass-sheet degeneracy can in principle be broken using magnification effect (see Broadhurst et al. 1995, 2005). In order to make full use of this method, the unlensed source counts at a given magnitude threshold must be known accurately. Given the photometric calibration uncertainties, which at the faint magnitudes one is usually dealing with easily amount to 0.1 mag, thus an

---

[*] Appendix A is only available in electronic form at
http://www.edpsciences.org

uncertainty of ~10% in the unlensed source counts is typical. As shown by Schneider et al. (2000), this level of uncertainty removes a great deal of the power of the magnification method to break the mass-sheet degeneracy.

Several attempts have been made recently to measure the cluster mass profiles with weak lensing. However, as shown in Clowe & Schneider (2001, 2002), it is extremely difficult to distinguish e.g. isothermal from NFW profiles at high significance using weak lensing data alone. The authors also conclude that these difficulties mostly arise as a consequence of the mass-sheet degeneracy transformation. Therefore additional information needs to be included, such as combining the weak lensing data with strong lensing (see e.g. Kneib et al. 2003; Smith et al. 2004). Another example was given by Sand et al. (2004) using combined strong lensing and stellar kinematics data of the dominating central galaxy. This approach offers valuable extra constraints, however a detailed strong-lens modelling is required (Bartelmann & Meneghetti 2004; Dalal & Keeton 2003).

In this paper we use a combined strong and weak lensing mass reconstruction to determine the mass and the mass distribution of clusters. We reconstruct the gravitational potential ψ, since it locally determines both the lensing distortion (for weak lensing) as well as the deflection (for strong lensing). The method extends the idea from Bartelmann et al. (1996) and Seitz et al. (1998). Its novel feature is that we directly include strong lensing information. Further, weak lensing reconstruction is extended to the critical parts of the cluster and we include individual redshift information of background sources as well as of the source(s) being multiply imaged. This allows us to break the mass-sheet degeneracy and accurately measure the cluster mass and mass distribution. The method is tested using simulations, and in Bradač et al. (2005, hereafter Paper II) we apply it to the cluster RX J1347.5−1145. In this paper we first briefly present the basics of gravitational lensing in Sect. 2. In Sect. 3 we give an outline of the reconstruction method (detailed calculations are given in the Appendix). We test the method using N-body simulations, and we present the results in Sect. 4. The conclusions and summary are the subject of Sect. 5.

2. Gravitational lensing preliminaries

Throughout this paper we follow the notation of Bartelmann & Schneider (2001), who give a detailed account of the gravitational lensing theory presented here.

We start by considering a lens having a projected surface mass density $\Sigma(\theta)$, where θ denotes the (angular) position in the lens plane. We define the dimensionless surface density $\kappa(\theta)$ for a fiducial source located at a redshift $z \to \infty$ and a lens (deflector) at $z = z_d$

$$\kappa(\theta) = \frac{\Sigma(\theta)}{\Sigma_{cr}} \quad \text{where} \quad \Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_\infty}{D_d D_{d,\infty}}. \tag{1}$$

D_∞, D_d, and $D_{d,\infty}$ are the angular diameter distances between the observer and the source at $z \to \infty$, the observer and the lens, and the lens and the source, respectively. The choice of scaling with Σ_{cr} is motivated by the fact that lenses with $\Sigma \geq \Sigma_{cr}$

(i.e. $\kappa \geq 1$) are strong enough to form multiple images. In this paper, however, *strong* lensing refers to multiple imaging only, while *weak* lensing means that the lensing effect is treated as statistical in nature, although it is also applied to the lens region with $\kappa > 1$, traditionally called the strong lensing regime.

We define the deflection potential $\psi(\theta)$

$$\psi(\theta) = \frac{1}{\pi} \int_{\mathfrak{R}^2} d^2\theta' \, \kappa(\theta') \, \ln|\theta - \theta'| \tag{2}$$

which is related to κ via the Poisson equation

$$\nabla^2 \psi(\theta) = 2\kappa(\theta). \tag{3}$$

Also the shear $\gamma = \gamma_1 + i\gamma_2$ and the deflection angle α are related to ψ, where

$$\gamma_1 = \frac{1}{2}(\psi_{,11} - \psi_{,22}), \quad \gamma_2 = \psi_{,12}, \quad \alpha = \nabla\psi. \tag{4}$$

The relations (3) and (4) are written for the source at redshift $z \to \infty$; we note, however, that they hold for any redshift. Since we will work with sources at different redshifts (both for strong as well as for weak lensing), we factorise the redshift dependence of the lens convergence κ, the shear γ, and the deflection angle α by

$$\kappa(\theta, z) = Z(z)\kappa(\theta), \quad \gamma(\theta, z) = Z(z)\gamma(\theta),$$
$$\alpha(\theta, z) = Z(z)\alpha(\theta). \tag{5}$$

$Z(z)$ is the so-called "cosmological weight" function:

$$Z(z) \equiv \frac{D_\infty D_{d,s}}{D_{d,\infty} D_s} H(z - z_d), \tag{6}$$

where $D_{d,s}$ and D_s, are the angular diameter distances between the lens and the source, and the observer and the source at a redshift z, respectively. $H(z - z_d)$ is the Heaviside step function and accounts for the fact that galaxies located at $z < z_d$ are not lensed[1].

In the case of weak lensing, the information on the lens potential is contained in the transformation between the source ellipticity $\epsilon^{(s)}$ and image ellipticity ϵ. It is given as a function of reduced shear $g(\theta, z)$ (see Seitz & Schneider 1997)

$$\epsilon^{(s)} = \begin{cases} \dfrac{\epsilon - g(\theta, z)}{1 - g^*(\theta, z)\epsilon} & \text{for } |g(\theta, z)| \leq 1, \\[3mm] \dfrac{1 - g(\theta, z)\epsilon^*}{\epsilon^* - g^*(\theta, z)} & \text{for } |g(\theta, z)| > 1 \end{cases} \tag{7}$$

where

$$g(\theta, z) = \frac{Z(z)\gamma(\theta)}{1 - Z(z)\kappa(\theta)}. \tag{8}$$

Galaxies are intrinsically elliptical, and therefore one cannot disentangle the effect of lensing from the intrinsic properties in Eq. (7) using a single galaxy image. Hence, the weak lensing effect needs to be treated in statistical sense. More precisely,

---

[1] To evaluate the angular diameter distances we assume the ΛCDM cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and Hubble constant $H_0 = 70$ km s^{-1} Mpc^{-1}.

we can Taylor expand the expression (7) (e.g. for the case of $|g(\boldsymbol{\theta}, z)| \le 1$) and recall that $\left|\epsilon^{(s)}\right| < 1$ by definition to obtain

$$
\begin{aligned}
\epsilon(z) &= \frac{\epsilon^{(s)} + g(\boldsymbol{\theta}, z)}{1 + g^*(\boldsymbol{\theta}, z)\epsilon^{(s)}} \\
&= \left(\epsilon^{(s)} + g(\boldsymbol{\theta}, z)\right) \sum_{k=0}^{\infty} (-1)^k (g^*(\boldsymbol{\theta}, z))^k (\epsilon^{(s)})^k.
\end{aligned}
\tag{9}
$$

A similar expansion can be obtained for the case $|g(\boldsymbol{\theta}, z)| > 1$. If we assume that the intrinsic ellipticity distribution has moments $\left\langle \epsilon^{(s)k} \right\rangle = 0$ for each k except $k = 0$ we get the known expression for the expectation value of the image ellipticity at redshift z

$$
\left\langle \epsilon(z) \right\rangle = \begin{cases} g(\boldsymbol{\theta}, z) & \text{if } \left|g(\boldsymbol{\theta}, z)\right| < 1, \\ \dfrac{1}{g^*(\boldsymbol{\theta}, z)} & \text{otherwise.} \end{cases}
\tag{10}
$$

This relation is particularly simple due to the convenient definition of ϵ. In the approximation $\kappa \ll 1$, $|\gamma| \ll 1$ (thus $|g| \ll 1$) the expectation value is given by $\left\langle \epsilon(z) \right\rangle = \gamma(\boldsymbol{\theta}, z)$.

3. The cluster mass reconstruction method

The idea of combining strong and weak lensing constraints is not new, it has been previously discussed by Abdelsalam et al. (1998), Kneib et al. (2003), Smith et al. (2004), and others. The method presented here, however, has some important differences. For example in Abdelsalam et al. (1998) the authors reconstruct the pixelized version of the surface mass density κ. A similar method for strong lensing constraints only has recently also been presented by Diego et al. (2004). We argue, however, that using the potential ψ is favourable, since κ, γ, and α locally depend upon the potential ψ – cf. Eqs. (3), (4) – and all can be quantified from the latter. γ and α, on the other hand, are non-local quantities of κ. In other words, the mass density on a finite field does not describe the shear and the deflection angle in this field. If a finite field is used, one usually employs Fourier analysis; in this case, γ in fact corresponds to original κ plus all its periodic continuations.

Further, even though not easy to implement, we decided to keep the parametrisation of the mass-distribution as general as possible. In Kneib et al. (2003) and Smith et al. (2004), on the other hand, the strong and weak lensing constraints were compared in a Bayesian approach in the form of simple, parametrised models. In addition, the weak lensing constraints were not used to the very centre of the cluster and redshifts of individual sources were not included.

3.1. The outline of the method

The main idea behind the method is to describe the cluster mass-distribution by a fully general lens, using the values of the deflection potential ψ on a regular grid. We then define a penalty function χ^2 and minimise it with respect to the values of ψ_k. The convergence κ, the shear γ, and the deflection angle α at an arbitrary position in the field are obtained by finite differencing and bilinear interpolation methods. The number of
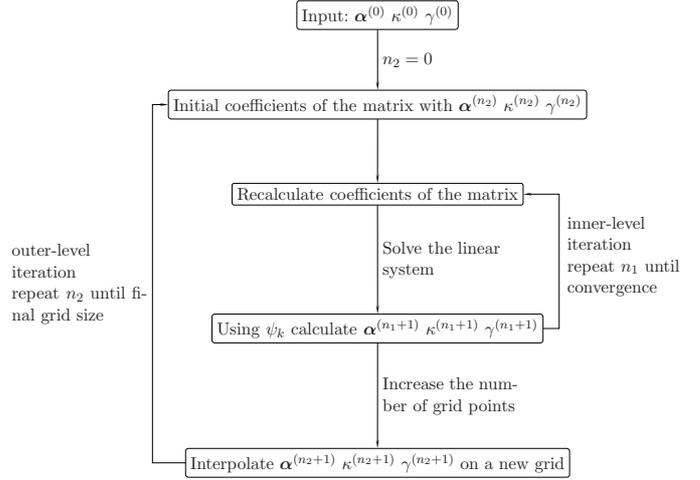

Fig. 1. The outline of the two-level iteration process.

grid points we use for ψ_k is $(N_x + 2) \times (N_y + 2)$; the extension by one row and one column at each side is needed to perform the finite differencing at each inner $N_x \times N_y$ grid point.

We define the χ^2-function as follows

$$
\chi^2(\psi_k) = \chi_\epsilon^2(\psi_k) + \chi_M^2(\psi_k) + \eta R(\psi_k).
\tag{11}
$$

$\chi_\epsilon^2(\psi_k)$ contains information from statistical weak lensing, whereas in $\chi_M^2(\psi_k)$ we include the multiple imaging properties. $R(\psi_k)$ is a regularisation term multiplied by the regularisation parameter η. The regularisation is a function of the potential and disfavours small-scale fluctuations in the surface mass density.

In order to find the minimum χ^2 solution, we solve

$$
\frac{\partial \chi^2(\psi_k)}{\partial \psi_k} = 0.
\tag{12}
$$

This is in general a non-linear set of equations, which we solve in an iterative manner. We linearise this system and in the first step we start from some trial solution, to calculate its non-linear terms (see the Appendix for details). We solve the corresponding equations and repeat this procedure until a convergence is achieved. Inverting the resulting matrix of $\sim \left(N_x \times N_y\right)^2$ elements for finding a solution of the linear system is difficult in general even for grids with a small number of cells. However, as it turns out, the resulting matrix is sparse and the system is therefore computationally inexpensive to solve.

The reconstruction is performed in a two-level iteration process, outlined in Fig. 1. We will refer to the iteration process mentioned above for solving the linear system of equations as inner-level, where steps n_1 are repeated until convergence of κ. The outer-level iteration is performed for the purpose of regularisation (as described in detail in Sect. 3.3). In order to penalise small-scale fluctuations in the surface mass density, we start the reconstruction with a coarse grid (large cell size). Then for each n_2 step we increase the number of grid points in the field and compare the new reconstructed $\kappa^{(n_2)}$ with the one from the previous iteration $\kappa^{(n_2-1)}$ (or with the initial input value $\kappa^{(0)}$ for $n_2 = 0$). The second-level iterations are performed until the final grid size is reached and convergence is achieved.
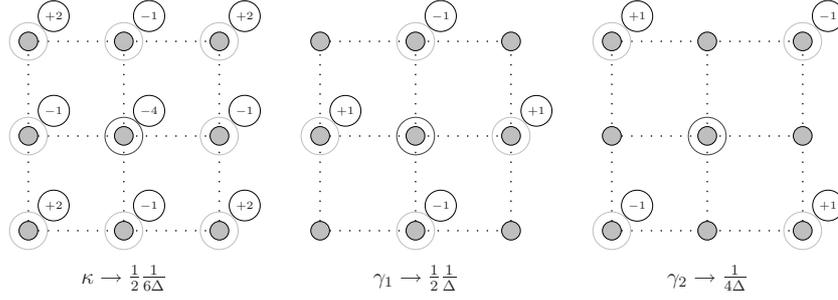
$$\kappa \rightarrow \tfrac{1}{2}\tfrac{1}{6\Delta} \qquad\qquad \gamma_1 \rightarrow \tfrac{1}{2}\tfrac{1}{\Delta} \qquad\qquad \gamma_2 \rightarrow \tfrac{1}{4\Delta}$$

Fig. 2. The finite differencing coefficients for κ (*left*), γ_1 (*middle*) and γ_2 (*right*). E.g. for κ we use a formula including 9 points, the multiplicative factor is given at the bottom, the individual coefficients in the circle. Thus for the middle point $(0, 0)$ we get $\kappa(0,0) = \tfrac{1}{2}\tfrac{1}{6\Delta}\Big[2\left[\psi(-1,1) + \psi(1,1) + \psi(-1,-1) + \psi(1,-1)\right] - \left[\psi(0,1) + \psi(-1,0) + \psi(1,0) + \psi(0,-1)\right] - 4\psi(0,0)\Big].$

3.2. Technical aspects

In this section we will briefly describe some technical aspects of how we calculate the lensing quantities κ, γ, and α at an arbitrary position within the field from the potential ψ_k on a grid.

We use the finite differencing method with 9 grid points to calculate κ, 5 points for γ, and 4 points for α (see Abramowitz & Stegun 1972). The coefficients used for κ and γ are given in Fig. 2, the case of α is discussed in Appendix A.2. To evaluate $\kappa(\theta)$, $\gamma(\theta)$, and $\alpha(\theta)$ at a position θ within the field, bilinear interpolation is used.

Note, that the dimensionality of the problem is not $(N_x + 2) \times (N_y + 2)$. Because the transformation $\psi(\theta) \rightarrow \psi(\theta) + \psi_0 + a \cdot \theta$ leaves κ and γ invariant, the potential needs to be fixed at three points (see Seitz et al. 1998; Bartelmann et al. 1996). These thus fix the constant and linear term in the invariance transformation. If this is not the case, a minimum in χ^2 would correspond to a three-dimensional subspace of possible solutions. The choice of the three points, and the corresponding values of the potential are arbitrary. Although the transformation $\psi(\theta) \rightarrow \psi(\theta) + a \cdot \theta$ changes the deflection angle α, it only causes a translation of the source plane, which is not an observable. Therefore, even in the presence of strong lensing, three points of the potential need to be held fixed.

The mass-sheet degeneracy transformation of the potential is given by $\psi \rightarrow \psi' = (1 - \lambda)\theta^2/2 + \lambda\psi$. However since we aim at lifting this degeneracy, in contrast to Seitz et al. (1998) the potential ψ_k, is *not* held fixed at an additional, fourth point. The dimensionality of the problem is thus $N_{\rm dim} = (N_x + 2)(N_y + 2) - 3$.

3.3. The χ^2-function

In this section we will describe contributions to the χ^2-function, starting with the statistical weak lensing.

For $N_{\rm g}$ galaxies with measured ellipticities ϵ_i we define the χ^2_ϵ as

$$\chi^2_\epsilon(\psi_k) = \sum_{i=1}^{N_{\rm g}} \frac{|\epsilon_i - \langle\epsilon\rangle|^2}{\sigma_i^2}, \tag{13}$$

where

$$\sigma_i^2 = \left(1 - |\langle\epsilon\rangle|^2\right)^2 \sigma_{\epsilon^{\rm s}}^2 + \sigma_{\rm err}^2. \tag{14}$$

Note that $\langle\epsilon\rangle$ refers to the expectation value of the ellipticity over redshift space, not to an ensemble average over different galaxies and is derived from Eq. (10).

In Bradač et al. (2004a)[2] we argue that χ^2_ϵ can give biased results for lenses for which many galaxies have $|g| \simeq 1$. It would be better to work with a log-likelihood function with a probability distribution that properly describes the distribution of observed ellipticities. Unfortunately, such an approach is not viable here (as will become obvious later on). However, in general clusters do not have a large fraction of galaxies with $|g| \simeq 1$ and we show in Bradač et al. (2004a) that for these lenses the χ^2-minimisation is sufficient.

One of the major strengths of this statistical weak lensing reconstruction technique is the possibility to simultaneously include constraints from multiple image systems to the weak lensing data in a relatively straightforward manner. The simplest approach to strong lensing is to perform the so-called "source plane" modelling; i.e. to minimise the projected source position difference. Consider a multiple image system with the source at redshift $z_{\rm s}$ and with $N_{\rm M}$ images located at θ_m. We define $b_m = \theta_m - Z(z_{\rm s})\alpha(\theta_m) - \beta_{\rm s}$ and the corresponding χ^2-function is given by

$$\chi^2_{\rm M} = \sum_{m=1}^{N_{\rm M}} b_m^{\rm T} \mathcal{S}^{-1} b_m, \tag{15}$$

where $\mathcal{S}$ is the covariance matrix. $\beta_{\rm s}$ is the average source position and for simplicity we calculate it using the deflection angle information from the previous iteration $n_1 - 1$. If the measurement errors on image positions are distributed isotropically, $\mathcal{S}$ reduces to a diagonal matrix given by $\mathcal{S} = {\rm diag}(\sigma_{\rm s,1}^2, \sigma_{\rm s,2}^2)$ with $\sigma_{\rm s,1}^2 = \sigma_{\rm s,2}^2$. $\sigma_{\rm s,1}$ and $\sigma_{\rm s,2}$ are the errors on image positions, projected onto the source plane. For simplicity, however, we do not perform a projection of the error ellipse from the image plane onto the source plane. Instead, we keep $\mathcal{S}$ constant throughout the reconstruction. Therefore we avoided the numerical problem of a diverging $\chi^2_{\rm M}$ function if one of the images lies at the position of the critical curve for the corresponding redshift.

We are aware of the fact that the approach we use is not optimal (see e.g. Kochanek 2004). If only multiple imaging is

---

[2] We would like to point out here that there is a typo in that paper. All $w_i = 1/\sigma_i$ in the text and plot labels need to be replaced by $w_i = 1/\sigma_i^2$, which has been used in the calculations.
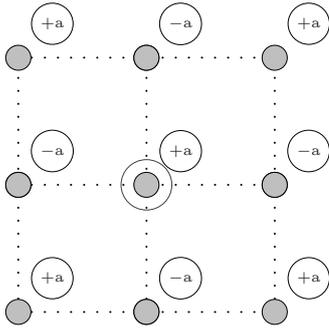
Fig. 3. The intrinsic invariance of the method. If we alternately add/subtract a constant a along the diagonals the shear γ does not change (cf. Fig. 2), but κ changes in the similar way with a constant now being $2a/3$.

used, the resulting best-fit model is biased towards high magnification factors, since errors on the source plane are magnified when projected back to the image plane (this information we do not use). In our case, however, the model also needs to take into account the constraints from statistical (weak) lensing and therefore the high magnification models are in fact discarded. In addition, if e.g. one considers the errors matrix in the image plane to be diagonal, the corresponding matrix in the source plane would have large off-diagonal terms. Throughout this paper we therefore consider the errors in the source plane to be isotropic, since this may in fact be a better approximation, as sources are on average more circular than their lensed images. In practice the location of the multiple images are usually known very accurately, leading to a very narrow minimum of $\chi^2_{\rm M}$ in the parameter space. In practice, multiple image constraints are satisfied nearly perfectly and exact values of errors on image positions are of lesser importance.

Since the minimisation of χ^2_ϵ can lead to a potential that reconstructs the noise in the data, the solution needs to be regularised. Even without measurement errors, the intrinsic ellipticities would still produce pronounced small-scale noise peaks in the final reconstruction. In addition, the method presented here has an intrinsic invariance if no multiple imaging information is used and the weak lensing approximation $g \simeq \gamma$ applies. Namely, we can alternately add/subtract a constant a along diagonals of the potential (chess-board like structure, as sketched in Fig. 3). This transformation would on the one hand not affect γ, but on the other it would cause a similar change (with a constant $2a/3$) in κ – compare with Fig. 2. Thus in the $|g| \ll 1$ regime, where $\langle \epsilon \rangle = g \simeq \gamma$ these stripes would show up in the resulting κ-map. This problem can, however, be very efficiently cured with regularisation.

Since we want to measure the cluster mass, the regularisation should not influence breaking the mass-sheet degeneracy. For example, one of the possibilities considered by Seitz et al. (1998) for regularisation function was $R = \sum_{i,j=1}^{N_x,N_y} |\nabla \kappa|^2$. However, as the authors mentioned, such regularisation would tend to flatten the profile and therefore affect the mass-sheet degeneracy breaking. Their maximum-entropy regularisation with moving prior (i.e. the prior in the regularisation is not kept constant, but adapted in the process of minimisation) does not

flatten the profile, however it is very cumbersome to express its derivative in linear terms of ψ_k. Motivated by the success of moving prior in maximum-entropy regularisation, we choose a very simple prescription for the regularisation function. We start off by a relatively coarse grid, since if the number of grid points $N_{\rm dim}$ is much smaller than the number of galaxies, the resulting reconstruction is not able to follow the noise pattern. In each second-level iteration step we gradually increase the number of grid points and compare the resulting κ map $\kappa^{(n_2)}$ with that from the previous iteration $\kappa^{(n_2-1)}$ linearly interpolated on a finer grid, thus

$$R = \sum_{i,j=1}^{N_x,N_y} \left(\kappa_{ij}^{(n_2)} - \kappa_{ij}^{(n_2-1)} \right)^2 . \tag{16}$$

For the case of $n_2 = 0$ we use an initial guess for κ which can in practice be obtained from strong lensing, direct finite-field reconstruction, parametrised model fitting to weak lensing data, or simply set to $\kappa_{ij}^{(0)} = 0$. This form of regularisation is relatively easy to implement and is in addition very efficient in removing the stripes (mentioned above) in the final reconstruction. If enough n_2 iteration steps are used this form of regularisation also does not substantially affect the ability to break the mass-sheet degeneracy, since the information of initial $\kappa^{(0)}$ is lost and different initial conditions do not bias the results (if the signal from lensing is high enough).

Finally a word on the regularisation constant η. This parameter should, in theory, be set such to ensure $\chi^2/N_{\rm d.o.f.} \sim 1$. In practice, however, it is difficult to determine its optimal value (in the critical lensing regime). As outlined in Geiger & Schneider (1998) the probability distribution of measured ellipticities is not a Gaussian and therefore the minimum value of χ^2 has no particular meaning. In practice, setting η such that the resulting $\chi^2/N_{\rm d.o.f.} \sim 1$ (where $N_{\rm d.o.f.}$ is in our case the number of galaxies used for strong and weak lensing) is valid is a good guess for this parameter. In addition, one adjusts η low enough for the method to have enough freedom to adapt to the information in the data and large enough for not allowing the solutions that follow the noise pattern. As a rule of thumb it is usually better to set η high and increase the number of iterations and hence allowing κ to change only slowly. Since the reconstruction is done in a two-level iteration and in addition multiple-image information is included, the method can successfully adapt to the data and the results are not very sensitive to the precise value of η. The resulting smoothness level of the mass maps should reflect the quality of data. The "smoothing scale" depends upon the combination of the grid size and regularisation. The final potential map should be void of any structures on scales smaller than the mean separation between galaxies used for weak lensing. We will shortly return to this point in Sect. 4.3.

3.4. Initial conditions

For the purpose of the regularisation, given by Eq. (16), we need to employ the initial conditions. In addition, if a realistic model is used for the initial conditions (and as trial solution), it is very helpful to break the internal degeneracy, i.e. to
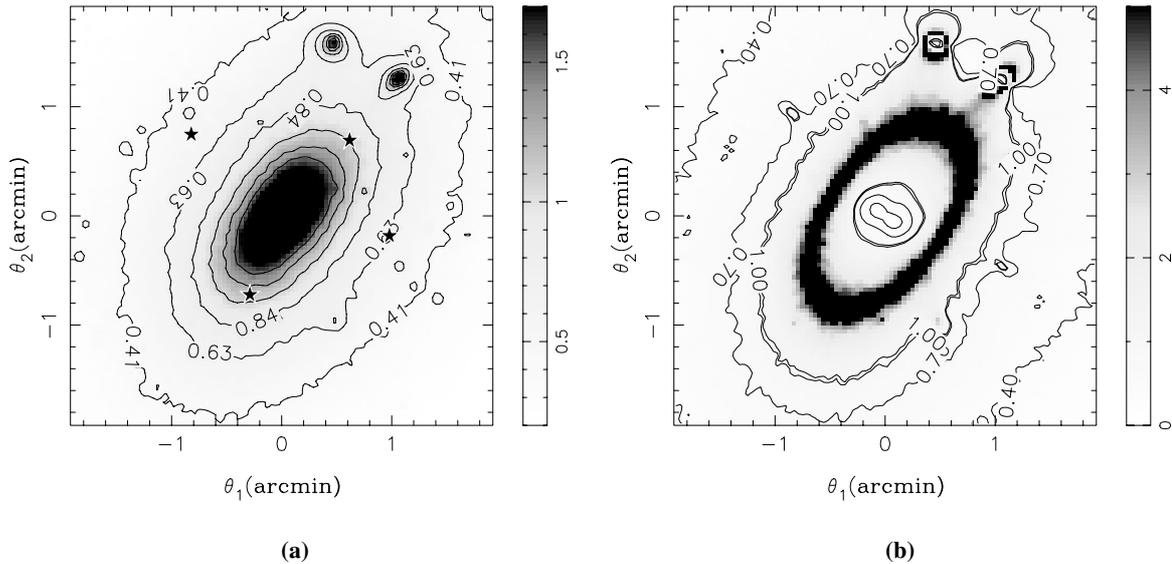
Fig. 4. The gravitational lensing properties of a simulated cluster used for generating mock catalogues for statistical weak lensing and for the multiple image system. **a)** The surface mass density κ; **b)** the absolute value of the reduced shear $|g|$, both for a source at $z_s \to \infty$, are given in gray-scale and contours. The stars in **a)** denote the image positions of a four-image system at $z_s = 1.76$ which we use for the reconstruction.

distinguish galaxies that have $|g| \leq 1$ from those with $|g| > 1$ in the first step, and thus allow for a faster convergence. Breaking this degeneracy is desirable, since otherwise the method has difficulties in "climbing" over the $|g| = 1$ region (especially if *no* multiple-image systems are included). Since we use individual redshifts of background galaxies we do not have well-defined critical curves (i.e. positions in the source plane where $|g| = 1$), as their position depends on the source redshift. In spite of this fact, the transition still poses a difficulty.

In our case different initial conditions are employed. For the initial model $\kappa^{(0)}$ we use three different scenarios: $\kappa^{(0)} = 0$ (and $\alpha^{(0)} = \mathbf{0}$, $\gamma^{(0)} = 0$) across the whole field (hereafter *I0*), $\kappa^{(0)}$ taken from the best fit non-singular isothermal ellipsoid model NIE for the multiple image system described in Sect. 4.2 (hereafter *IM*) and a non-singular isothermal sphere model NIS with scaling and core radius being the same as in *IM* (hereafter *IC*). The same models are used also to obtain the initial coefficients of the linear system (see Appendix) (for *I0* we use $\gamma_{1,2} = 0$). These different initial models help us to explore the effects of regularisation and the capability of the reconstruction method to adapt to the data.

4. Cluster mass reconstruction from simulated data

4.1. Mock catalogues

We generate mock catalogues using a cluster from the high-resolution *N*-body simulation by Springel et al. (2001). The cluster is taken from the S4 simulation (for details see the aforementioned paper) and was simulated in the framework of the ΛCDM cosmology with density parameters $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$, shape parameter $\Gamma = 0.21$, the normalisation of the power spectrum $\sigma_8 = 0.9$, and Hubble constant $H_0 = 70$ km s^{-1} Mpc^{-1}. The cluster simulation consists of almost 20 million particles, each with a mass of 4.68×10^7 $M_\odot$

and a gravitational softening length of 0.7 h^{-1} kpc. Due to the high mass resolution, the surface mass density κ-map can be obtained by directly projecting the particles (in our case along the *z*-axis) onto a 1024^2 grid (of a side length 4 Mpc) using the NGP (nearest gridpoint) assignment.

In what follows we try to generate the weak and strong lensing data to resemble as close as possible the data on the cluster RX J1347.5−1145 we will use in Paper II. The surface mass density of the cluster is therefore scaled to have a sizeable region where multiple imaging is possible within a 3.8×3.8 arcmin^2 field, for sources at redshifts $z \gtrsim 1$ and a cluster at $z_d = 0.4$. The Einstein radius for a fiducial source at $z \to \infty$ is roughly $\theta_E \sim 1'$, giving a line-of-sight integrated mass within this radius of $\sim 5 \times 10^{14}$ $M_\odot$. The cut-out of the resulting κ map (for $z \to \infty$, thus $Z(z) = 1$) we use can be seen in Fig. 4a.

The lensing properties are calculated as described in detail in Bradač et al. (2004b). The Poisson equation for the lens potential ψ – cf. Eq. (3) – is solved on the grid in Fourier space with a DFT (Discrete Fourier Transformation) method using the FFTW library written by Frigo & Johnson (1998). The two components of the shear $\gamma_{1,2}$ and the deflection angle α are obtained by finite differencing methods applied to the potential ψ. These data are then used to generate the weak lensing catalogues as well as the multiple image systems. The absolute value of the reduced shear (again for a source with $z \to \infty$) is shown in Fig. 4b.

The weak lensing data are obtained by placing N_g galaxies on a 3.8×3.8 arcmin^2 field. We have simulated two different catalogues, one with $N_g = 148$ galaxies with positions corresponding to those from *R*-band weak lensing data of the cluster RX J1347.5−1145 and one with $N_g = 210$ galaxies corresponding to the *I*-band data used in Paper II. In this way we simulate the effects of "holes", resulting from cluster obscuration and bright stars in the field.

The intrinsic ellipticities ϵ_s are drawn from a Gaussian distribution, each component is characterised by $\sigma = \sigma_{\epsilon^s} = 0.2$. We use the same redshifts as those measured in the R and I-band data, respectively. The catalogues have average redshifts for background sources of $\langle z_I \rangle = 1.18$ and $\langle z_R \rangle = 1.14$. The corresponding cosmological weights are evaluated assuming the ΛCDM cosmology (the same parameters are used as for the cluster simulations).

The measurement errors ϵ^{err} on the observed ellipticities are drawn from a Gaussian distribution with $\sigma = \sigma_{err} = 0.1$ (each component) and added to the lensed ellipticities. We considered also the measurement errors on the redshifts of the galaxies to simulate the use of photometric redshifts. These have $\sigma_{zerr} = 0.1 (1 + z_i)$ (see Bolzonella et al. 2000); in adding the errors we ensured that the resulting redshifts are always positive. We have also simulated the presence of outliers in the redshift distribution, 10% of our background sources (chosen at random) are considered outliers, for these we randomly add/subtract $\Delta z = 1$ to their redshifts (which already include random errors). The lensed ellipticities are obtained using Eq. (9) and interpolating the quantities κ, and γ at the galaxy position using bilinear interpolation, considering the redshifts including errors. In contrast, for the purpose of reconstruction we then consider galaxies to be at their "original" redshifts (thus equal to the observed redshifts in the data of RX J1347.5−1145).

4.2. Multiple imaging

To obtain a four-image system from the simulation we use the method described in detail in Bradač et al. (2004b). With the MNEWT routine from Press et al. (1992) we solve the lens equation for a given source position inside the asteroid caustic. The source is assumed at a redshift of $z_s = 1.76$. Once we have the image positions, their magnifications are calculated and the fifth image is eliminated, since it is usually too dim and would not be observable.

The errors on image positions can be conservatively estimated (for the data we use in Paper II) to $\sim 0\!''\!.3$. Since we need errors in the source plane, we set them by a factor of five smaller (in agreement with the average magnification factor for this system), i.e. $\sigma_{s,m} = 0\!''\!.06$ for both coordinates (see discussion in Sect. 3.3).

We also use this system to obtain one of the $\kappa^{(0)}$ models, needed for the purpose of strong and weak lensing reconstruction. We perform image plane minimisation and fit an NIE model (Kormann et al. 1994), given by

$$\kappa(\theta') = \frac{b_0}{2 \sqrt{\frac{1+|\epsilon_g|}{1-|\epsilon_g|} \left(r^2_{c,nis} + (\theta'_1)^2 \right) + (\theta'_2)^2}}, \qquad (17)$$

where θ' is calculated w.r.t. the semi-major axis of the cluster surface mass density. We allow the scaling b_0, ellipticity $|\epsilon_g|$ and position angle ϕ_g to vary. The best fit model for this system has values of $\{b_0, |\epsilon_g|, \phi_g\} = \{0\!''\!.97, 0.30, 1.01\}$. We fix the core radius r_c to $0\!''\!.1$. For model fitting we use C-minuit (James & Roos 1975), a routine which is part of the CERN Program Library.

Table 1. Reconstructed cluster mass within a cylinder of 340 h^{-1} kpc radius around the cluster centre from simulations of mock catalogues resembling I-band (*left*) and R-band (*right*) weak lensing data and one 4-image system. Three different initial conditions are used. We use the best fit model from the multiple image system *IM* (see Sect. 4.2), the *IC* model (NIS with same scaling and core radius as *IM*) and *I0* with $\kappa^{(0)} = 0$ on all grid points. In the last line the input mass M_s from the simulation is given. The variance of the mass estimate is given and in brackets we give for comparison the velocity dispersion of an SIS having the same enclosed mass within 340 h^{-1} kpc.

	$M_{I,s}$ [$10^{15} M_\odot$]	[$\sigma_{I,SIS}$] [km s^{-1}]	$M_{R,s}$ [$10^{15} M_\odot$]	[$\sigma_{R,SIS}$] [km s^{-1}]
IM	1.07 ± 0.02	[1720]	1.04 ± 0.02	[1710]
IC	1.04 ± 0.02	[1700]	1.00 ± 0.02	[1670]
I0	0.88 ± 0.03	[1560]	0.82 ± 0.03	[1510]
$M_s(<340\ h^{-1}\ \mathrm{kpc})$			0.99	[1670]

4.3. Weak lensing mass-reconstruction using simulated data

The mock catalogues are used to test the performance of the reconstruction method. We obtain the solution of the linear system with the UMFPACK routine for solving asymmetric sparse linear systems (Davis & Duff 1999). We perform 30 second-level iterations, each time increasing N_x and N_y by one, starting with an initial 20×20 grid (since we use regularisation from the first iteration, we do not need to start the reconstruction with a very course grid, where the resulting matrix is not sparse and a different routine should be applied to find a good solution). The results of the reconstructions are shown in Fig. 5 for $N_g = 210$ galaxies distributed in the same manner as the I-band data and $N_g = 148$ galaxies distributed as the R-band data of the cluster described in Paper II. The initial regularisation parameter was set to $\eta = 400$ for I-band $\eta = 200$ for R-band data and adapted in each step to ensure $\chi^2/N_{d.o.f.} \sim 1$. Two different initial values for the regularisation parameter are used since the numbers of galaxies in the two bands are different. It is very comforting to observe that the reconstructed maps do not depend crucially on the initial κ-model we use.

From the reconstructed maps we estimate the mass within the $1\!'\!.5$ radius from the centre of the cluster (for a redshift $z_d = 0.4$ this corresponds to 340 h^{-1} kpc) projected along the line-of-sight. For this purpose we generate 10 mock catalogues for each band and did the reconstruction again with the three different initial models. We list the resulting average mass obtained from the catalogues in Table 1 for both the I- and R-band mock catalogues. All the mass estimates are similar; note, however, that the galaxy catalogues following the I- and R-band data have galaxies partly in common and the errors are therefore correlated. We find the enclosed mass of the simulated cluster to be $(1.0 \pm 0.1) \times 10^{15} M_\odot$, which is very close to the input value of $M_s(<340\ h^{-1}\ \mathrm{kpc}) = 0.99 \times 10^{15} M_\odot$. The 1-$\sigma$ error is estimated from the variance of mass determinations from different mock catalogues.

The results show that our method is *effectively* able to break the mass-sheet degeneracy and is, as a consequence, very
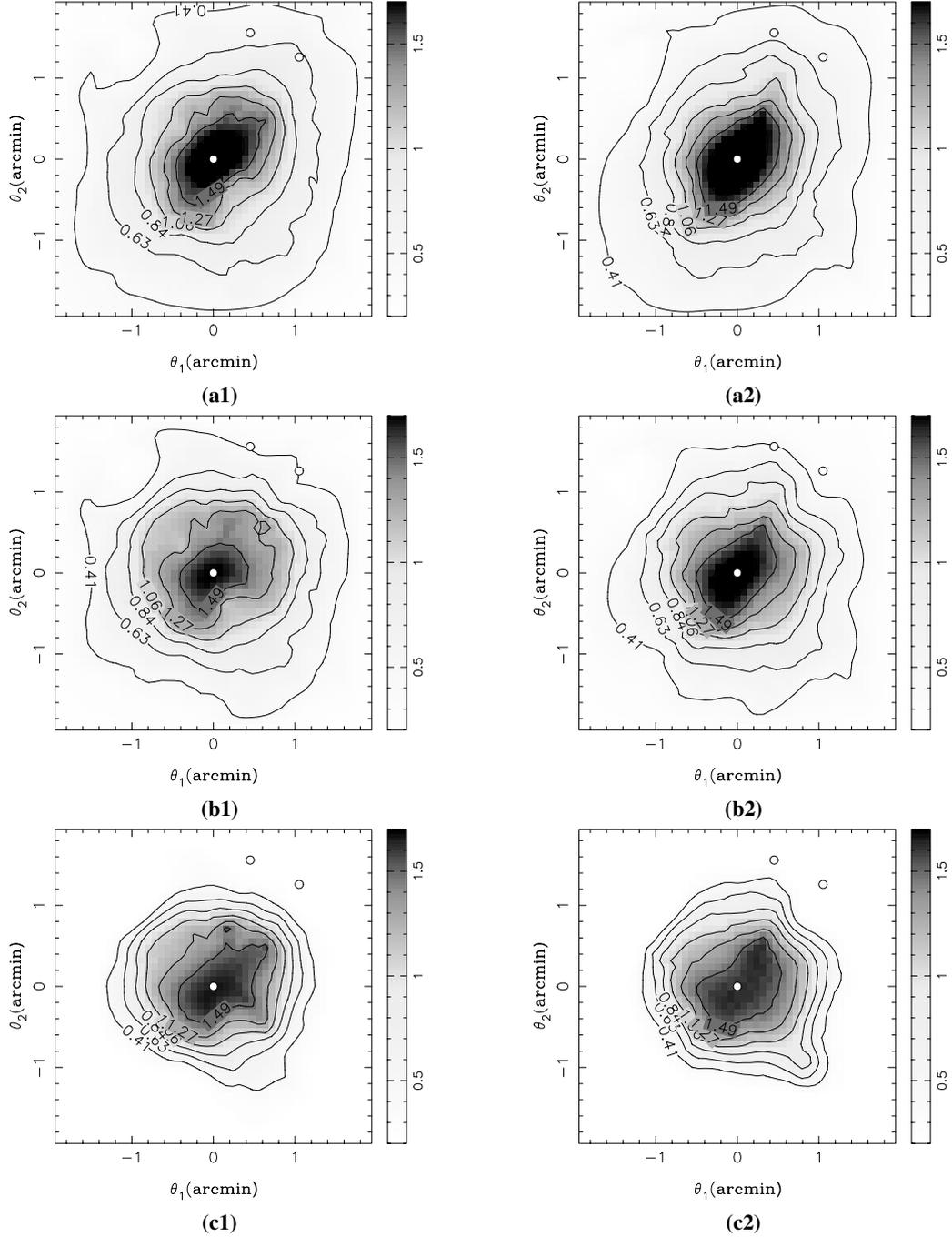
(a1)

(a2)

(b1)

(b2)

(c1)

(c2)

Fig. 5. κ-maps obtained from combined strong and weak lensing reconstruction of the simulated data. *Left panels* show the reconstructions using $N_g = 210$ galaxies distributed in the same manner as the *I*-band, while for the *right panels* we use $N_g = 148$ galaxies distributed in the same manner as the *R*-band weak lensing data for RX J1347.5−1145 (see Paper II). The galaxies have been lensed by an *N*-body simulated cluster. Different initial conditions are used for the reconstruction. In **a1)**–**a2)** we use best fit model from the multiple image system *IM* (see Sect. 4.2) in **b1)**–**b2)** we use the *IC* model, an NIS model with the same scaling and core radius as *IM* and in **c1)**–**c2)** we use *I0*, i.e. $\kappa^{(0)} = 0$ on all grid points. The positions of the cluster centre and two major subclumps are plotted as white circles.

efficient in reproducing the cluster mass also at radii signifi-
cantly larger than the Einstein radius of the cluster. It is also
very encouraging that the results are nearly independent of
the initial $\kappa^{(0)}$ used for the regularisation. Note that a single
multiple-image system does not by itself break this degener-
acy, we would need at least two different redshift multiple im-
age systems to break the mass-sheet degeneracy with strong

lensing data alone. In such a case the strong lensing gives con-
straints on the mass enclosed within the Einstein radius for a
given source redshift and since the critical curves depend on
the source redshift, we can constrain mass at two different radii
and the degeneracy is broken. The combiniation of weak and
strong lensing is the more powerful, the more different the red-
shift of the source of the multiple images is from the median

redshift of the galaxies from which the weak lensing measurements are obtained, and the less symmetric the arrangement of multiple images is w.r.t. the center of the cluster.

Unfortunately we can not resolve both clumps present in the simulations. This is due to the fact that the number density of background sources is low and the internal smoothing scale (i.e. the average distance between two source galaxies) is large; with a number density of ~100 arcmin^{-2} the clumps can be easily resolved.

We have also performed additional reconstructions in which we multiplied the original values of κ of the simulated cluster by 0.75 and 1.25. This enables us to confirm that the agreement between input mass and reconstructed mass is not just accidental. We have generated new multiple image systems and new mock catalogues as before. We do not, however, perform a new strong-lensing reconstruction, for $\kappa^{(0)}$ we intentionally use the same (i.e. in this case "wrong") initial conditions as before. The old *IM* model would not fit the image positions any longer, since they have changed with the scaling of κ. The reconstructed masses of the increased κ simulation are in good agreement with the input values. The differences between different models are comparable (slightly smaller) to the ones shown in Table 1. For the lower κ simulation, the reconstructed values are on average the same as the input value, however the scatter is larger. This is expected, since the lens in this case is weaker and the breaking of the mass-sheet degeneracy is difficult in this case (with the quality of data used here).

As an additional test we also consider a redshift distribution with $\langle z \rangle = 1.6$ for the sources used for weak lensing and regenerate the mock catalogues. The accuracy of the determination of the enclosed mass increases. However, more importantly, we also better reconstruct the shape of the mass distribution, since high-redshift galaxies (when their shape is measured reliably) contribute most to the signal and improve the accuracy of the reconstruction.

5. Conclusions

In this paper we develop a new method based on Bartelmann et al. (1996) to perform a combined weak and strong lensing cluster mass reconstruction. The particular strength of this method is that we extend the weak lensing analysis to the critical parts of the cluster. In turn, this enables us to directly include multiple imaging information to the reconstruction. Strong and weak lensing reconstructions are performed on the same scales, in contrast to similar methods proposed in the past where weak lensing information on much larger radii than the Einstein radius θ_E was combined with strong lensing information (see e.g. Kneib et al. 2003).

We test the performance of the method on simulated data and conclude that if a quadruply imaged system combined with weak lensing data and individual photometric redshifts is used, the method can very successfully reconstruct the cluster mass distribution. With a relatively low number density of background galaxies, 15 arcmin^{-2}, we are effectively able to reproduce the main properties of the simulated cluster. In addition, with larger number densities ~100 arcmin^{-2} of background sources, accessible by HST, the substructures in the cluster can be resolved and the mass determination further improved.

We determine the enclosed mass within 340 h^{-1} kpc of the simulated cluster to be $(1.0\pm0.1)\times10^{15}\ M_\odot$, which is very close to the input value of $M_s(<340\ h^{-1}$ kpc$) = 0.99 \times 10^{15}\ M_\odot$. We have shown, that with the data quality we use we are effectively able to break the mass-sheet degeneracy and therefore obtain the mass and mass-distribution estimates without prior assumptions on the lensing potential.

In addition, the reconstruction algorithm can be improved in many ways. First, we use for the multiply imaged system only the information of the image positions. The reconstruction method can, however, be modified to include the morphological information of each extended source. Instead of using a regular grid, one would have to use adaptive grids and decrease the cell sizes around each of these images. This will be a subject of a future work. Second, the photometric redshift determination does not only give the most likely redshift given the magnitudes in different filters, but also the probability distribution for the redshift. This information can be included in the reconstruction. In addition, source galaxies without redshift information can be included and different regularisation schemes can be considered.

Finally, the slight dependence on the initial conditions is getting weaker the higher the number density of background galaxies and/or multiple image systems are. In addition, it is of advantage to have a large spread in the redshift efficiency factors Z of the background galaxies. For example, deep ACS images of clusters with a usable number density of $n \sim 120$ arcmin^{-2}, or future observations with the James Webb Space Telescope will most likely make the dependence on the initial conditions negligible.

In Paper II we will show the application of this method on the cluster RX J1347.5−1145 and confirm that a combination of strong and weak lensing offers a unique tool to pin down the masses of galaxy-clusters as well as their mass distributions.

References

Abdelsalam, H. M., Saha, P., & Williams, L. L. R. 1998, AJ, 116, 1541
Abramowitz, M., & Stegun, I. A. 1972, Handbook of Mathematical Functions, Handbook of Mathematical Functions (New York: Dover)
Allen, S. W. 1998, MNRAS, 296, 392
Bartelmann, M., & Meneghetti, M. 2004, A&A, 418, 413
Bartelmann, M., Narayan, R., Seitz, S., & Schneider, P. 1996, ApJ, 464, L115

Bartelmann, M., & Schneider, P. 2001, Phys. Rep., 340, 291

Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, A&A, 363, 476

Bradač, M., Lombardi, M., & Schneider, P. 2004a, A&A, 424, 13

Bradač, M., Schneider, P., Lombardi, M., et al. 2004b, A&A, 423, 797

Bradač, M., Erben, T., Schneider, P., et al. 2005, 437, 49

Broadhurst, T., Takada, M., Umetsu, K., et al. 2005, ApJ, 619, L143

Broadhurst, T. J., Taylor, A. N., & Peacock, J. A. 1995, ApJ, 438, 49

Clowe, D., & Schneider, P. 2001, A&A, 379, 384

Clowe, D., & Schneider, P. 2002, A&A, 395, 385

Dalal, N., & Keeton, C. R. 2003 [arXiv:astro-ph/0312072]

Davis, T. A., & Duff, I. S. 1999, ACM Trans. Math. Softw., 25, 1

Diego, J. M., Protopapas, P., Sandvik, H. B., & Tegmark, M. 2004 [arXiv:astro-ph/0408418]

Ettori, S., & Lombardi, M. 2003, A&A, 398, L5

Frigo, M., & Johnson, S. 1998, in 1998 ICASSP Conf. Proc., 3, 1381

Gavazzi, R., Mellier, Y., Fort, B., Cuillandre, J.-C., & Dantel-Fort, M. 2004, A&A, 422, 407

Geiger, B., & Schneider, P. 1998, MNRAS, 295, 497

James, F., & Roos, M. 1975, Computer Phys. Commun., 10, 343

Kaiser, N., & Squires, G. 1993, ApJ, 404, 441

Kneib, J., Hudelot, P., Ellis, R. S., et al. 2003, ApJ, 598, 804

Kochanek, C. S. 2004 [arXiv:astro-ph/0407232]

Kormann, R., Schneider, P., & Bartelmann, M. 1994, A&A, 284, 285

Lombardi, M., Rosati, P., Blakeslee, J. P., et al. 2005 [arXiv:astro-ph/0501150]

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in C. The art of scientific computing (Cambridge: University Press), 2nd ed.

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. 1997, A&A, 318, 687

Seitz, S., Schneider, P., & Bartelmann, M. 1998, A&A, 337, 325

Smith, G. P., Kneib, J., Smail, I., et al. 2004 [arXiv:astro-ph/0403588]

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726

Tyson, J. A., Wenk, R. A., & Valdes, F. 1990, ApJ, 349, L1

Online Material

Appendix A: The linear problem for ψ_k

In this section we present details of the method outlined in Sect. 3.

We aim to solve the equation

$$\frac{\partial \chi^2_\epsilon(\psi_k)}{\partial \psi_k} + \frac{\partial \chi^2_M(\psi_k)}{\partial \psi_k} + \eta \frac{\partial R(\psi_k)}{\partial \psi_k} = 0. \tag{A.1}$$

This is in general a non-linear system of equations. We try to solve it in an iterative way by linearising the equation in terms of ψ_k and keeping the non-linear terms fixed at each iteration step. The resulting system is then written in the form

$$\mathcal{B}_{jk}\psi_k = V_j, \tag{A.2}$$

where the matrix $\mathcal{B}_{jk}$ and vector V_j contain the contributions from the non-linear part. In the following sections we will describe the contributions to Eq. (A.1) in turn.

A.1. The weak lensing analysis

The χ^2_ϵ for the weak lensing case is given in Eq. (13). From now on we consider in detail only the $|g| \leq 1$ case; for $|g| > 1$, the calculations are done in the same fashion. First we plug into Eq. (13) the expectation value of observed ellipticities (i.e. the reduced shear g) obtaining

$$\chi^2_\epsilon(\psi_k) = \sum_{i=1}^{N_g} \frac{\left|\epsilon - \frac{Z\gamma}{1-Z\kappa}\right|^2}{\sigma^2} = \sum_{i=1}^{N_g} \frac{|\epsilon - Z\epsilon\kappa - Z\gamma|^2}{(1-Z\kappa)^2 \sigma^2}, \tag{A.3}$$

where κ, γ, and σ depend on θ_i only and Z depends on the redshift of the ith source. Note that for simplicity we omit the index i to ϵ, κ, γ, σ and Z, although these quantities are different for every galaxy. As described in Sect. 3.2 using finite differencing and bilinear interpolation, we can write κ and γ at each galaxy position as a linear combination of ψ_k. This is expressed in the following matrix notation

$$\gamma_1(\theta_i) = \mathcal{G}^{(1)}_{ik}\psi_k, \quad \gamma_2(\theta_i) = \mathcal{G}^{(2)}_{ik}\psi_k, \quad \kappa(\theta_i) = \mathcal{K}_{ik}\psi_k, \tag{A.4}$$

where the matrices $\mathcal{G}^{(1)}_{ik}$, $\mathcal{G}^{(2)}_{ik}$, and $\mathcal{K}_{ik}$ are composed of numerical factors described in Sect. 3.2. Now we consider the denominator of Eq. (A.3) $\hat{\sigma}^2_\leq = (1 - Z\kappa)^2 \sigma^2$ fixed at each iteration step (the subscript $\leq$ denotes the $|g| \leq 1$ case) and differentiate the following term of Eq. (A.3)

$$\frac{1}{2}\frac{\partial}{\partial \psi_k}\frac{|\epsilon - Z\epsilon\kappa - Z\gamma|^2}{\hat{\sigma}^2_\leq} =$$

$$-\frac{Z}{\hat{\sigma}^2_\leq}\left[(\epsilon_1 - Z\epsilon_1\kappa - Z\gamma_1)\left(\epsilon_1 \frac{\partial \kappa}{\partial \psi_k} + \frac{\partial \gamma_1}{\partial \psi_k}\right)\right.$$

$$\left. + (\epsilon_2 - Z\epsilon_2\kappa - Z\gamma_2)\left(\epsilon_2 \frac{\partial \kappa}{\partial \psi_k} + \frac{\partial \gamma_2}{\partial \psi_k}\right)\right] =$$

$$\frac{Z^2}{\hat{\sigma}^2_\leq}\left[\mathcal{G}^{(1)}_{ij}\mathcal{G}^{(1)}_{ik} + \mathcal{G}^{(2)}_{ij}\mathcal{G}^{(2)}_{ik} + \epsilon_1\left(\mathcal{G}^{(1)}_{ij}\mathcal{K}_{ik} + \mathcal{K}_{ij}\mathcal{G}^{(1)}_{ik}\right)\right.$$

$$\left. + \epsilon_2\left(\mathcal{G}^{(2)}_{ij}\mathcal{K}_{ik} + \mathcal{K}_{ij}\mathcal{G}^{(2)}_{ik}\right) + \left(\epsilon_1^2 + \epsilon_2^2\right)\mathcal{K}_{ij}\mathcal{K}_{ik}\right]\psi_k$$

$$-\frac{Z^2}{\hat{\sigma}^2_\leq}\left[\epsilon_1\mathcal{G}^{(1)}_{ij} + \epsilon_2\mathcal{G}^{(2)}_{ij} + \left(\epsilon_1^2 + \epsilon_2^2\right)\mathcal{K}_{ij}\right] \tag{A.5}$$

where ϵ_1 and ϵ_2 are the two components of the measured ellipticity of galaxy i (again omitting the index and we divided Eq. (A.1) by two for simplicity). We sum over all galaxies used for the weak lensing analysis and obtain a linear problem for ψ_k at each iteration step. The same approach can be used for the $|g| > 1$ case, where $\hat{\sigma}^2_>$ is kept constant and is given by $\hat{\sigma}^2_> = Z_i^2 |\gamma|^2 \sigma_i^2$.

A.2. The strong-lensing term

Following the prescription from the previous section we now write the deflection angle in a matrix form

$$\alpha_1(\theta_m) = \mathcal{D}^{(1)}_{ik}\psi_k, \quad \alpha_2(\theta_m) = \mathcal{D}^{(2)}_{ik}\psi_k. \tag{A.6}$$

Both matrices give the finite differencing form for the gradient of the potential, in particular we use the central differencing formula, i.e. $\alpha_1(0,0) = \frac{1}{2\Delta}(\psi(1,0) - \psi(-1,0))$ and $\alpha_2(0,0) = \frac{1}{2\Delta}(\psi(0,1) - \psi(0,-1))$.

The χ^2 contribution to strong lensing is given in Eq. (15). The source position β_s is kept constant at every iteration step, and is evaluated using the deflection angle information $\alpha^{(n_1-1)}$ from the previous iteration

$$\beta_s = \frac{1}{N_M}\sum_{m=1}^{N_M}\left(\theta_m - Z(z_s)\alpha^{(n_1-1)}(\theta_m)\right). \tag{A.7}$$

We differentiate the following term in χ^2_M (for the x_1-coordinate)

$$\frac{1}{2}\frac{\partial}{\partial \psi_k}\frac{(\theta_{m,1} - Z(z_s)\alpha_1(\theta_m) - \beta_{s,1})^2}{\sigma^2_{s,m(1)}} =$$

$$-\frac{(\theta_{m,1} - \beta_{s,1})\mathcal{D}^{(1)}_{ij} - Z(z_s)\mathcal{D}^{(1)}_{ij}\mathcal{D}^{(1)}_{ik}\psi_k}{\sigma^2_{s,m(1)}}. \tag{A.8}$$

The expression for the x_2-coordinate is obtained by exchanging $1 \rightarrow 2$. After summation of both terms over all images m we get a set of equations which are linear in ψ_k and can be readily included in Eq. (A.2). In principle we could also use $\alpha(\psi_k)$ in Eq. (A.7), the expression $\partial\chi^2_M/\partial\psi_k$ would remain linear in ψ_k. However since the first approach is computationally much simpler, we use the former.

A.3. The final result

In the previous section we describe how we linearise the contributions of weak and strong lensing, now we can write the coefficients in the Eq. (A.2). Note that the contribution of the regularisation term (with χ^2-contribution given in Eq. (16)) is

already linear in ψ_k and therefore the full matrix $\mathcal{B}_{jk}$ is given in the form

$$
\begin{aligned}
\mathcal{B}_{jk} = \sum_{i=1}^{N_{\text{gal}}} & \Big[a_{11}(i)\mathcal{G}_{ij}^{(1)}\mathcal{G}_{ik}^{(1)} + a_{22}(i)\mathcal{G}_{ij}^{(2)}\mathcal{G}_{ik}^{(2)} \\
& + a_{13}(i)\left(\mathcal{G}_{ij}^{(1)}\mathcal{K}_{ik} + \mathcal{G}_{ik}^{(1)}\mathcal{K}_{ij}\right) \\
& + a_{23}(i)\left(\mathcal{G}_{ij}^{(2)}\mathcal{K}_{ik} + \mathcal{G}_{ik}^{(2)}\mathcal{K}_{ij}\right) + a_{33}(i)\left(\mathcal{K}_{ij}\mathcal{K}_{ik}\right) \Big] \\
& + \sum_{m=1}^{N_{\text{M}}} b_{11}(m)\mathcal{D}_{mj}^{(1)}\mathcal{D}_{mk}^{(1)} + b_{22}(m)\mathcal{D}_{mj}^{(2)}\mathcal{D}_{mk}^{(2)} \\
& + \eta \sum_{g} \mathcal{K}_{gj}\mathcal{K}_{gk},
\end{aligned}
\tag{A.9}
$$

where the sums over i, g and m denote summation over all galaxies with ellipticity measurements, all grid points, and all images in the multiple imaged system, respectively. The V_j vector carries the information of all constant terms in Eq. (A.1)

$$
\begin{aligned}
V_j = \sum_{i=1}^{N_{\text{gal}}} & a_1(i)\mathcal{G}_{ij}^{(1)} + a_2(i)\,\mathcal{G}_{ij}^{(2)} + a_3(i)\,\mathcal{K}_{ij} \\
& + \eta \sum_{g} \kappa^{(n_2-1)}\mathcal{K}_{gj} \\
& + \sum_{m=1}^{N_{\text{M}}} b_1(m)\mathcal{D}_{mj}^{(1)} + b_2(m)\,\mathcal{D}_{mj}^{(2)}.
\end{aligned}
\tag{A.10}
$$

The coefficients a now differ depending on whether we are in the $|g| \leq 1$ or $|g| > 1$ regime. For $|g| \leq 1$ the coefficients are given by

$$
a_{11}(i) = a_{22}(i) = \frac{Z^2}{\hat{\sigma}_{\leq}^2}; \quad a_{13}(i) = \frac{Z^2}{\hat{\sigma}_{\leq}^2}\epsilon_1;
$$

$$
a_{23}(i) = \frac{Z^2}{\hat{\sigma}_{\leq}^2}\epsilon_2; \quad a_{33}(i) = \frac{Z^2}{\hat{\sigma}_{\leq}^2}|\epsilon|^2;
$$

$$
a_1(i) = \frac{Z}{\hat{\sigma}_{\leq}^2}\epsilon_1; \; a_2(i) = \frac{Z}{\hat{\sigma}_{\leq}^2}\epsilon_2; \; a_3(i) = \frac{Z}{\hat{\sigma}_{\leq}^2}|\epsilon|^2.
\tag{A.11}
$$

For $|g| > 1$ case we get

$$
a_{11}(i) = a_{22}(i) = \frac{Z^2}{\hat{\sigma}_{>}^2}|\epsilon|^2; \quad a_{13}(i) = \frac{Z^2}{\hat{\sigma}_{>}^2}\epsilon_1;
$$

$$
a_{23}(i) = \frac{Z^2}{\hat{\sigma}_{>}^2}\epsilon_2; \quad a_{33}(i) = \frac{Z^2}{\hat{\sigma}_{>}^2};
$$

$$
a_1(i) = \frac{Z}{\hat{\sigma}_{>}^2}\epsilon_1; \; a_2(i) = \frac{Z}{\hat{\sigma}_{>}^2}\epsilon_2; \; a_3(i) = \frac{Z}{\hat{\sigma}_{>}^2}.
\tag{A.12}
$$

The coefficients b are carrying the information about the multiple imaged system:

$$
b_{11}(m) = \frac{Z^2(z_{\text{s}})}{\sigma_{\text{s},m(1)}^2}; \quad b_{22}(m) = \frac{Z^2(z_{\text{s}})}{\sigma_{\text{s},m(2)}^2};
$$

$$
b_1(m) = \frac{Z(z_{\text{s}})\,(\theta_{m,1} - \beta_{\text{s},1})}{\sigma_{\text{s},m(1)}^2};
$$

$$
b_2(m) = \frac{Z(z_{\text{s}})\,(\theta_{m,2} - \beta_{\text{s},2})}{\sigma_{\text{s},m(2)}^2},
\tag{A.13}
$$

where $\theta_{m,1}$, $\theta_{m,2}$, and $\sigma_{\text{s},m(1,2)}$ are defined in Sect. 3.3. For the reasons mentioned in Sect. 3.3 we consider the measurement errors projected to the source plane isotropic, we set $\sigma_{\text{s},m(1,2)}$ equal for the purpose of the reconstruction.