

High resolution weak lensing mass mapping combining shear and flexion[★]

F. Lanusse^{1,2}, J.-L. Starck², A. Leonard³, and S. Pires²

¹ McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
e-mail: flanusse@andrew.cmu.edu

² Laboratoire AIM, UMR CEA-CNRS-Paris, Irfu, SAp, CEA Saclay, 91191 Gif-sur-Yvette Cedex, France

³ Department of Physics and Astronomy, University College London, Gower Place, London WC1E 6BT, UK

Received 9 February 2016 / Accepted 12 March 2016

ABSTRACT

Aims. We propose a new mass mapping algorithm, specifically designed to recover small-scale information from a combination of gravitational shear and flexion. Including flexion allows us to supplement the shear on small scales in order to increase the sensitivity to substructures and the overall resolution of the convergence map without relying on strong lensing constraints.

Methods. To preserve all available small scale information, we avoid any binning of the irregularly sampled input shear and flexion fields and treat the mass mapping problem as a general ill-posed inverse problem, which is regularised using a robust multi-scale wavelet sparsity prior. The resulting algorithm incorporates redshift, reduced shear, and reduced flexion measurements for individual galaxies and is made highly efficient by the use of fast Fourier estimators.

Results. We tested our reconstruction method on a set of realistic weak lensing simulations corresponding to typical HST/ACS cluster observations and demonstrate our ability to recover substructures with the inclusion of flexion, which are otherwise lost if only shear information is used. In particular, we can detect substructures on the 15'' scale well outside of the critical region of the clusters. In addition, flexion also helps to constrain the shape of the central regions of the main dark matter halos.

Key words. gravitational lensing: weak – methods: data analysis – galaxies: clusters: general – dark matter

1. Introduction

Probing the mass distribution of the various structural components of the Universe is a fundamental tool for constraining cosmological models. Gravitational lensing, which is sensitive to the total mass content, is particularly well suited to this task. One of the strengths of gravitational lensing is its ability to probe the dark matter content from the largest cosmological scales down to the galaxy scale, providing valuable constraints on all intermediary scales.

In particular, on the galaxy cluster scale, gravitational lensing is an established tool for measuring the total mass distribution and investigating the nature of dark matter (e.g. Clowe et al. 2006; Massey et al. 2015). While the combination of gravitational shear and strong lensing constraints has proven very successful to resolve the fine structure of galaxy clusters (Bradač et al. 2005; Cacciato et al. 2006; Merten et al. 2009), strong lensing is only effective within the Einstein radius of the cluster, which represents for most clusters only a fraction of their total area. Outside of this region, constraints from gravitational shear are generally very poor below the arcminute scale, which makes the detection of substructures extremely unlikely. A promising avenue to bridge the gap between shear and strong lensing and to help resolve substructures outside of the innermost regions of galaxy clusters is gravitational flexion, which has already been shown to provide valuable constraints on these intermediate scales (Leonard et al. 2007).

Measuring flexion has proven to be a difficult task in practice (Viola et al. 2012; Rowe et al. 2013). Originally, flexion estimators were derived from shapelet coefficients (Goldberg & Bacon 2005; Bacon et al. 2006). Another approach relied on moments of galaxy images (Okura et al. 2007; Goldberg & Leonard 2007; Schneider & Er 2008). More recently, a different approach to flexion estimation, called the analytic image method (AIM), has been proposed by Cain et al. (2011) based on forward-fitting analytical galaxy image models. Among the benefits of this promising new method are a complete invariance under the mass-sheet degeneracy and a better characterisation of the measurement errors.

Very few methods have been developed for mass mapping that includes flexion. The first reconstruction technique using flexion has been proposed in Bacon et al. (2006) as an extension of the Kaiser-Squires Fourier estimator (Kaiser & Squires 1993). A parametric reconstruction of Abell 1689 was attempted in Leonard et al. (2007) and Leonard et al. (2009), Leonard & King (2010) later proposed an extension of aperture mass filters to the case of flexion. Joint reconstruction based on maximum likelihood methods combining shear, flexion and strong lensing have also been proposed by Er et al. (2010) and most recently in Cain et al. (2016).

Recovering high resolution mass-maps from weak lensing is a difficult problem for two main reasons: the irregular sampling of the lensing field and the low signal-to-noise ratio (S/N) on small scales. While the lensing equations can be most easily inverted when the field is fully sampled on a regular grid, for instance using a Kaiser-Squires estimator, in practice the shear and flexion fields are only sampled at the position of the background

[★] Our mass mapping software, called Glimpse2D, is made freely available at <http://www.cosmostat.org/software/glimpse>

galaxies, which are not regularly distributed. In that case the lensing equations are no longer directly invertible and mass mapping becomes an ill-posed inverse problem, i.e., the solution is not unique and/or extremely unstable to measurement noise. To avoid this problem, a usual approach is to bin or smooth the input shear or flexion on a regular grid, which not only serves as a regularisation of the inverse problem but also acts as a denoising step by suppressing smaller scales, where the noise typically dominates. Although this reasoning holds for shear alone, adding flexion information greatly increases the S/N of structures on small scales, in which case binning or smoothing the input data can destroy relevant information.

The aim of this paper is to propose a new mass mapping methodology that is specifically designed to recover small-scale information from weak lensing by avoiding any binning or smoothing of the input data and supplementing shear information with flexion to increase the sensitivity to substructures. We address the mass mapping problem as a general ill-posed inverse problem which we regularise using a wavelet-based sparsity prior on the recovered convergence map. Similar sparse regularisation methods have already been proposed in the context of weak lensing mass mapping in Pires et al. (2009) for inpainting missing data and Leonard et al. (2012, 2014) for 3D mass mapping. In this work, the sparse regularisation framework not only allows us to solve the lensing equations even at very low sampling rates but it also provides a very robust multi-scale noise regularisation. Furthermore, despite the irregular sampling of the data, our algorithm relies on fast Fourier estimators to invert the lensing equations, by treating the computation of the Fourier transform (which is ill-posed in this case) as part of the same inverse problem. This makes our approach far more efficient than the costly finite differencing schemes usually used in maximum likelihood algorithms. Finally, the proposed algorithm accounts for reduced shear and flexion and incorporates individual redshift measurements of background sources.

This paper is structured as follows. We present a short overview of the weak lensing formalism and motivate the interest of flexion in Sect. 2. In Sect. 3 we introduce and demonstrate the effectiveness of our sparse regularisation approach in the simplified setting of reconstructing the convergence from shear alone. Then, in Sect. 4, we incorporate the non-linearity induced by the reduced shear, we add redshift information for individual galaxies and we extend the framework to include flexion. Finally, in Sect. 5 we demonstrate on realistic cluster simulations that our method is successful at reconstructing the surface mass density and illustrate the impact of the additional flexion information for the recovery of small scale substructures.

2. Weak gravitational lensing formalism

2.1. Weak lensing in the linear regime

Because gravitational lensing is sensitive to the total matter content of structures, it is an ideal probe to map the dark matter distribution. The presence of a massive lens along the line of sight will induce a distortion of the background galaxy images which can be described by a coordinate transformation (Bartelmann 2010) between the unlensed coordinates β and the observed image coordinates θ :

$$\beta = \theta - \nabla\psi(\theta), \quad (1)$$

where ψ is the 2D lensing potential encoding the deflection of light rays by the gravitational lens. The lensing potential can

be seen as generated from a source term, through the following Poisson equation:

$$\Delta\psi(\theta) = 2\kappa(\theta), \quad (2)$$

where κ is the convergence, a dimensionless surface density. If we consider a single thin lens, κ can be derived from the surface mass density of the lens Σ as:

$$\kappa(\theta) = \frac{\Sigma(\theta)}{\Sigma_{\text{critic}}}, \quad (3)$$

where Σ_{critic} is the critical surface density defined by a ratio of distances between the observer, the lens and the source:

$$\Sigma_{\text{critic}} = \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}}. \quad (4)$$

In this expression, D_S , D_L and D_{LS} are respectively the angular diameter distances to the source, the lens and between the lens and the source.

If the lensing effect is weak enough, the coordinate transformation can be approximated using a Taylor expansion to first order of Eq. (1) as:

$$\beta_i \simeq A_{ij}\theta_j, \quad (5)$$

where \mathbf{A} is the amplification matrix, defined in terms of the derivatives of the lensing potential and which takes the form:

$$A_{ij}(\theta) = \frac{\partial\beta_i}{\partial\theta_j} = \delta_{ij} - \partial_i\partial_j\psi(\theta), \quad (6)$$

$$\mathbf{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (7)$$

where γ is the complex shear, causing anisotropic distortions whereas κ only produces an isotropic magnification effect. However, in practice, when using galaxy ellipticities as a measurement of the lensing effect, neither the convergence nor the shear are directly observable. Indeed, the net contribution of weak lensing to the ellipticity of a galaxy is not the shear γ but the reduced shear g defined as $g = \frac{\gamma}{1-\kappa}$. The amplification matrix can equivalently be expressed as a function of the reduced shear by factorising a term $(1-\kappa)$ which yields:

$$\mathbf{A} = (1-\kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix}. \quad (8)$$

Note that in the weak lensing regime, for $\kappa \ll 1$, the reduced shear can be approximated to the actual shear, $\gamma \simeq g$. However, this assumption no longer holds for the purpose of mapping galaxy clusters where it is critical to correctly account for the reduced shear.

2.2. Gravitational flexion

The previous expressions for the coordinate transformations hold if the convergence and shear fields are assumed to be constant on the scale of the observed source images. When this assumption is not verified, the expansion of the coordinate transformation can be pushed to second order (Goldberg & Bacon 2005):

$$\beta_i \simeq A_{ij}\theta_j + \frac{1}{2}D_{ijk}\theta_j\theta_k, \quad (9)$$

where the third order lensing tensor D_{ijk} can be derived from $D_{ijk} = \partial_k A_{ij}$. This additional term gives rise to third order derivatives of the lensing potential which can be summarised as a spin-1 field $\mathcal{F} = \nabla\kappa$ and a spin-3 field $\mathcal{G} = \nabla\gamma$ respectively called first and second flexion.

Similarly to the case of the reduced shear, these two flexion components are not directly observable. Instead, only the reduced flexion fields are measured from galaxy images:

$$F = \frac{\mathcal{F}}{1 - \kappa}; \quad G = \frac{\mathcal{G}}{1 - \kappa}. \quad (10)$$

Furthermore, estimating in practice the second flexion G is extremely delicate and measurements are generally noise dominated and discarded. As a result, in the rest of this work, we will only include measurements of the reduced first flexion F .

Throughout this work, we assume reduced shear and flexion measurement provided by the AIM method of [Cain et al. \(2011\)](#). In particular, we will use $\sigma_F = 0.029 \text{ arcsec}^{-1}$ for the dispersion of flexion measurements, as reported by [Cain et al. \(2011\)](#) on HST data.

2.3. mass mapping using Fourier estimators

The most well known inversion technique for weak lensing mass mapping is the Kaiser-Squires inversion ([Kaiser & Squires 1993](#)) which consists in applying a minimum variance filter in Fourier space to both components of the shear, yielding one estimate of the convergence field with minimum noise variance, given by

$$\hat{\kappa}_\gamma = \frac{k_1^2 - k_2^2}{k^2} \tilde{\gamma}_1 + \frac{2k_1 k_2}{k^2} \tilde{\gamma}_2, \quad (11)$$

where $k^2 = k_1^2 + k_2^2$. This estimator is only valid for $k^2 \neq 0$ and as a result the mean value of the field cannot be recovered, which corresponds to the well-known mass sheet degeneracy. Assuming white Gaussian noise of variance σ_γ^2 on the input shear field, the Kaiser-Squires estimator features a simple flat noise power spectrum:

$$\langle \hat{\kappa}_\gamma^* \hat{\kappa}_\gamma \rangle = \sigma_\gamma^2. \quad (12)$$

Although this inversion technique has some very desirable properties including linearity and minimum variance, it has one major shortcoming, namely that is not well defined on bounded domains, even less so on domains with complex geometries and missing data, which occurs in actual surveys. Some alternatives to the simple Kaiser-Squires have also been developed to mitigate some of its shortcomings ([Pires et al. 2009](#); [Deriaz et al. 2012](#)) using iterative schemes.

Similarly to the Kaiser-Squires inversion for the shear, [Bacon et al. \(2006\)](#) proposed a minimum variance filter for estimating the convergence from first flexion \mathcal{F} . As the flexion is only the gradient of the convergence, these two quantities can be related in the Fourier domain as:

$$\tilde{\mathcal{F}}_1 = -ik_1 \tilde{\kappa}; \quad \tilde{\mathcal{F}}_2 = -ik_2 \tilde{\kappa}. \quad (13)$$

With the same approach as the Kaiser-Squires inversion, the two components of the first flexion can be combined in order to yield a minimum variance estimator of the convergence, which takes the simple expression:

$$\hat{\kappa}_F = \frac{ik_1 \tilde{\mathcal{F}}_1 + ik_2 \tilde{\mathcal{F}}_2}{k^2}. \quad (14)$$

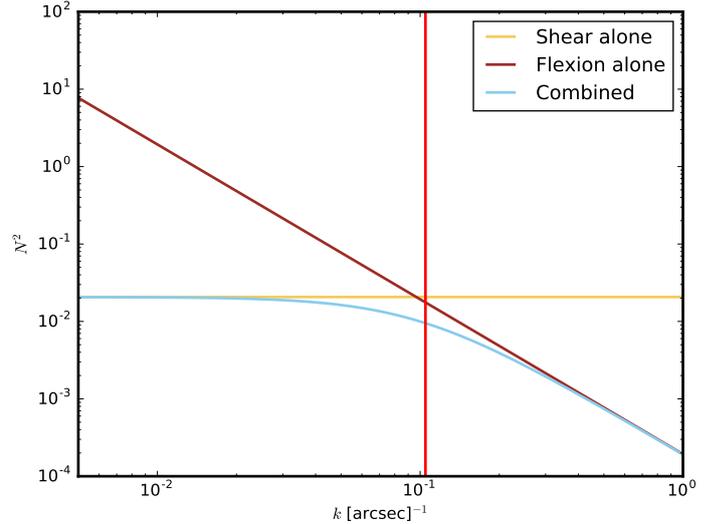


Fig. 1. Noise power spectra of the minimum variance estimators for shear alone, flexion alone and shear and flexion combined. The vertical red line indicates the arcminute scale. These noise power spectra assume $\sigma_F = 0.029 \text{ arcsec}^{-1}$, $\sigma_\gamma = 0.3$ and $n_g = 80 \text{ gal/arcmin}^2$ with 14 arcsec pixels.

Note that, just as with the Kaiser-Squires inversion, this expression is only valid for $k_1 \neq 0$ and $k_2 \neq 0$, which means that the flexion reconstruction is still subject to the mass sheet degeneracy. Contrary to the previous Kaiser-Squires inversion, this estimator does not have a flat noise power spectrum. Indeed, assuming uncorrelated flexion measurements with intrinsic variance σ_F^2 , the variance of this estimator is:

$$\langle \hat{\kappa}_F^* \hat{\kappa}_F \rangle = \frac{k_1^2}{k^4} \sigma_F^2 + \frac{k_2^2}{k^4} \sigma_F^2, \quad (15)$$

$$= \frac{1}{k^2} \sigma_F^2. \quad (16)$$

This makes the reconstruction of mass maps from flexion alone using this estimator very problematic on large scales where the noise will always dominate the signal. This means that applying any low-pass smoothing actually reduces the S/N of the reconstruction.

Figure 1 shows the noise power spectra of a shear inversion alone and flexion inversion alone, the arcminute scale is marked by the vertical line. As can be seen, the flexion noise power spectrum is much more favourable than the shear on small scales. However, it is only worth considering adding flexion information if the scale at which these two power spectra cross is still relevant. This scale only depends on the ratio of shear and flexion intrinsic dispersion and we find it to be around 65 arcsec for $\sigma_F = 0.029 \text{ arcsec}^{-1}$ and $\sigma_\gamma = 0.3$. As a result, based solely on considerations about the relative noise power spectra using these Fourier estimators, we argue that flexion can be used to help resolve sub-arcmin structures.

It is clear from [Fig. 1](#) that although flexion can help constrain small features, it is not competitive with respect to the shear on scales larger than 1 arcmin. To simultaneously benefit from both shear and flexion, [Bacon et al. \(2006\)](#) proposed a minimum variance filter combining both measurements. Although they present reconstructions using this combined filter, they do not explicitly provide its expression, which can easily be derived and takes the

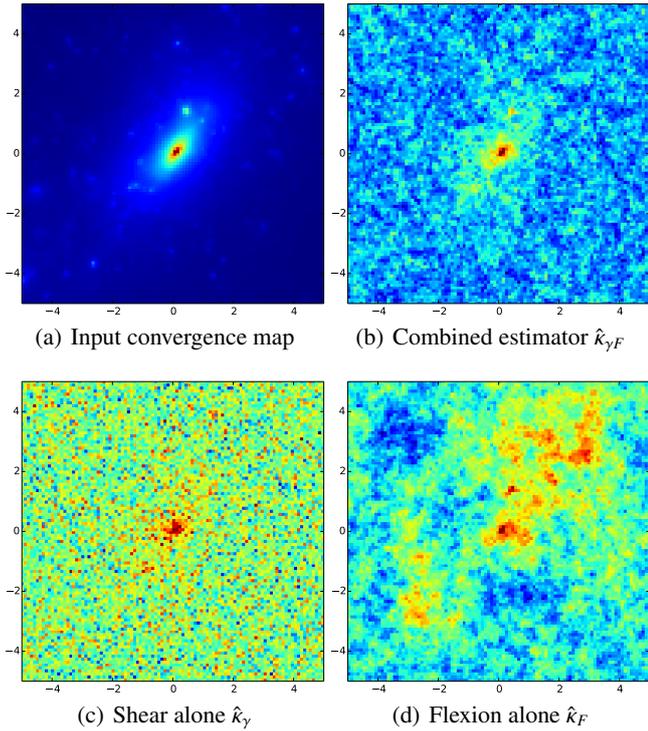


Fig. 2. *Top:* simulated 10×10 arcmin² field containing a massive cluster **a**) and its reconstruction using the combined minimum variance estimator **b**). *Bottom:* reconstructions using the shear alone Kaiser-Squires estimator **c**) and the flexion alone minimum variance estimator **d**). The noise level correspond to $\sigma_F = 0.04$ arcsec⁻¹, $\sigma_\gamma = 0.3$ and $n_g = 50$ gal/arcmin² with 6 arcsecs pixels.

form:

$$\hat{\kappa}_{\gamma F} = \frac{1}{k^2 + \frac{\sigma_F^2}{\sigma_\gamma^2}} \left(ik_1 \tilde{\mathcal{F}}_1 + ik_2 \tilde{\mathcal{F}}_2 + \frac{\sigma_F^2}{\sigma_\gamma^2} \left(\frac{k_1^2 - k_2^2}{k^2} \tilde{\gamma}_1 + \frac{2k_1 k_2}{k^2} \tilde{\gamma}_2 \right) \right). \quad (17)$$

The noise variance of this estimator is now:

$$\langle \hat{\kappa}_{\gamma F}^* \hat{\kappa}_{\gamma F} \rangle = \frac{\sigma_F^2}{k^2 + \frac{\sigma_F^2}{\sigma_\gamma^2}}. \quad (18)$$

As can be seen, on large scales, i.e. for small k we recover asymptotically the flat shear noise power spectrum while on small scales we recover the flexion estimator noise power spectrum. Figure 1 illustrates the noise power spectrum of this combined estimator (blue line) in realistic conditions. We see that the combined estimator starts improving over the shear alone around the arcminute scale.

An example of reconstruction using this combined estimator is shown in Fig. 2. The input convergence map is reconstructed from shear alone, flexion alone and combining shear and flexion. As expected, the shear alone reconstruction is very noisy, especially on small scales because of its flat noise power spectrum. On the contrary, the flexion alone reconstruction is noise dominated on scales larger than 1 arcmin. However, by combining both shear and flexion information, the noise is effectively suppressed on small scales and not amplified on large scales which makes the cluster and some of its substructure clearly identifiable without any additional filtering. Note however that this simple

example only illustrates the noise properties of these estimators but is unrealistic in the sense that it does not consider the problem of missing data.

These Fourier estimators are only valid in the weak lensing regime and should only be applied when the convergence is weak (when $\kappa \ll 1$) such that the observed reduced shear and flexion can be approximated to the true shear and flexion:

$$g = \frac{\gamma}{1 - \kappa} \simeq \gamma \quad \text{and} \quad F = \frac{\mathcal{F}}{1 - \kappa} \simeq \mathcal{F}. \quad (19)$$

Of course this assumption no longer holds at the close vicinity of galaxy clusters and it is therefore important to properly account for reduced shear and flexion for the purpose of mapping galaxy clusters. In those situations, κ is no longer negligible and the reconstruction problem becomes non-linear. As was noted in Seitz & Schneider (1995), it is still possible to recover the convergence using the previous linear estimators by iteratively solving the inversion problem, using at each iteration the previous estimate of the convergence to correct the measured reduced shear:

$$\kappa_{n+1}(\boldsymbol{\theta}) - \kappa_0 = \frac{1}{\pi} \int d\boldsymbol{\theta}' (1 - \kappa_n(\boldsymbol{\theta}')) \Re [D^*(\boldsymbol{\theta} - \boldsymbol{\theta}') g(\boldsymbol{\theta}')], \quad (20)$$

where $D(\boldsymbol{\theta})$ is the Kaiser-Squires convolution kernel and the constant κ_0 accounts for the fact that the constant of the convergence is not constrained by the data. This iterative process is generally found to converge quickly to the solution. Of course, a similar procedure can be implemented to correct the reduced flexion and we will use the same principle in the reconstruction algorithm presented in the next section.

2.4. Mass-sheet degeneracy and redshift dependency

One of the most notorious issues in weak lensing mass mapping is the so-called mass-sheet degeneracy (Seitz & Schneider 1996). This degeneracy is due to the fact that the shear is left invariant by the addition of a constant mass-sheet to the lens surface density. As a result, the observed reduced shear is invariant under the following λ -transformation:

$$\kappa' = \lambda\kappa + (1 - \lambda). \quad (21)$$

Indeed, if g' is the reduced shear generated by κ' then:

$$g' = \frac{\lambda\gamma}{1 - \lambda\kappa - 1 + \lambda} = \frac{\lambda\gamma}{\lambda - \lambda\kappa} = g. \quad (22)$$

By the same mechanism, the reduced flexion F is also invariant under the same λ -transformation: adding the flexion information will not affect the mass-sheet degeneracy.

This issue is particularly problematic for measuring the mass of galaxy clusters from weak-lensing. Indeed, these measurements are typically performed on small fields where it cannot simply be assumed that the convergence goes to zero outside the field without biasing the measurement (Bartelmann 1995).

However, although it is true that from shear alone the mass-sheet degeneracy can not be lifted, it can nonetheless be mitigated when including additional information about the relative distances between the lens and the different sources. Indeed, as was pointed out in Bradac et al. (2004), knowledge of individual photometric redshifts of background galaxies is enough to constrain the mass-sheet for strong enough lenses.

Let us consider $\gamma_\infty(\theta)$ and $\kappa_\infty(\theta)$ the shear and convergence of a given lens at redshift z_l , for sources at infinite redshift. The

actual shear and convergence applied to a galaxy at a specific redshift z_s can be expressed as $\kappa(\theta, z_s) = Z(z_s)\kappa_\infty(\theta)$ and $\gamma(\theta, z_s) = Z(z_s)\gamma_\infty(\theta)$ where Z is a cosmological weight, function of the redshifts of the source and lens, which can be expressed as a ratio of two critical surface densities:

$$Z(z_s) = \frac{\Sigma_{\text{crit}}^\infty}{\Sigma_{\text{crit}}(z_s)} H(z_s - z_l), \quad (23)$$

where H is the Heaviside step function, which only accounts for the fact that sources located at lower redshift than the lens are not lensed, and $\Sigma_{\text{crit}}^\infty = \lim_{z \rightarrow \infty} \Sigma_{\text{crit}}(z)$ is the critical surface density for sources at infinite redshift.

If we now consider the reduced shear measured on a galaxy at a specific redshift z_s as a function of κ_∞ and γ_∞ , it takes the form:

$$g(\theta, z_s) = \frac{Z(z_s)\gamma_\infty(\theta)}{1 - Z(z_s)\kappa_\infty(\theta)}. \quad (24)$$

This expression alone is just a rewriting of the measured reduced shear, which makes explicit the dependency on the redshift of the source. Therefore, the convergence is still subject to the mass-sheet degeneracy which can now be made explicit in term of z_s :

$$\kappa' = \lambda\kappa + \frac{1 - \lambda}{Z(z_s)}. \quad (25)$$

It is easily shown that such a transformation leaves the measured reduced shear invariant. However, it is specific to a given source redshift z_s and the reduced shear measured on a galaxy at a different redshift $z'_s \neq z_s$ is not invariant under the same transformation. As a result, simultaneously constraining the convergence to fit the reduced shear at different redshifts formally leads to $\lambda = 1$ and breaks the degeneracy. In practice however, this mechanism is only effective at lifting the degeneracy in regions where the convergence is strong enough. Indeed, in the linear regime the transformation $\kappa' = \kappa + \lambda$ has no dependency on the redshift of the source. Therefore, the mass-sheet degeneracy can only be lifted in the neighbourhood of strong enough lenses, where the weak lensing approximation no longer holds.

For the rest of this work, we will assume knowledge of the redshift z_L of the lens as well as individual photometric redshifts for the sources. For a source with a given photo- z distribution $p_i(z)$, we define the following lensing weight Z_i relating the convergence of the source and the convergence at infinite redshift:

$$Z_i = \int_{z_L}^{\infty} Z(z)p_i(z)dz = \int_{z_L}^{\infty} \frac{\Sigma_{\text{crit}}^\infty}{\Sigma_{\text{crit}}(z)} p_i(z)dz, \quad (26)$$

where $\Sigma_{\text{crit}}(z)$ is the critical density for a lens at redshift z_L and a source at redshift $z_s = z$ while $\Sigma_{\text{crit}}^\infty$ is the critical density for sources at infinite redshift. With this definition, the reduced shear g_i and reduced flexion F_i for a given source can be described as a function of the shear, flexion and convergence for sources at infinite redshift γ_∞ , F_∞ and κ_∞ :

$$g_i = \frac{Z_i\gamma_\infty}{1 - Z_i\kappa_\infty}; \quad F_i = \frac{Z_iF_\infty}{1 - Z_i\kappa_\infty}. \quad (27)$$

Finally, for convenience in the rest of the paper, we introduce the diagonal matrix $\mathbf{Z} = \text{diag}(Z_1, Z_2, \dots, Z_N)$, with N the number of galaxies in the survey.

3. A new mass mapping methodology

3.1. Dealing with irregularly sampled data

The global inversion methods presented in the previous section all assume knowledge of the shear on a grid. However, the shear is only sampled in practice at randomly distributed galaxy positions. Therefore a first step in all of these methods is to bin the shear catalogue in order to have a sufficient number of galaxies per pixel prior to performing the inversion. Even so, some regions are bound to remain empty because of various masks applied to the data. The simplest and most common procedure to handle missing data consists in applying a large Gaussian smoothing to the binned shear maps with a kernel larger than the typical size of the masks. Obviously, this strategy is far from optimal as it irremediably destroys small scale information.

A more sophisticated approach was proposed in Pires et al. (2009), based on sparse inpainting. This method regularises the inversion in the case of missing data by imposing on the solution a sparsity prior in a discrete cosine transform (DCT) dictionary. In effect, the resulting inpainted map features minimal B-modes leaking and unbiased power spectrum and bispectrum with respect to the original map. This method has been employed to map galaxy clusters in Jullo et al. (2014) where the authors advocate using a very fine binning of the data in order to have on average one galaxy per pixel.

We propose in this section a new method for reconstructing the convergence map, based on Fourier estimators, which does not require any binning or smoothing of the input shear. Although this approach allows us to preserve small-scale information which would otherwise be lost, it does turn the reconstruction problem into an ill-posed inverse problem, which we will address using sparse regularisation

3.1.1. Non-equispaced discrete fourier transform

When using Fourier based methods, the fundamental motivation behind binning or smoothing is to regularise the discrete fourier transform (DFT) of the shear, which is not well defined for data sampled on an irregular grid and which is known in this case as a non-equispaced discrete fourier transform (NDFT). The NDFT is of course no longer an orthogonal transform and is usually not even invertible. We will consider the case where only the spatial nodes $\mathbf{x} = (x_l)_{0 \leq l < M}$ are arbitrary while the frequency nodes $\mathbf{k} = \llbracket 0, N \rrbracket$ are N regularly spaced integers. Computing the NDFT from a set of Fourier coefficients $\hat{\mathbf{f}} = (\hat{f}_k)_{0 \leq k < N}$ simply amounts to evaluating the trigonometric polynomials:

$$\forall l \in \llbracket 0, M \rrbracket, \quad f_l = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{f}_k e^{2\pi i k x_l}. \quad (28)$$

This operation can more conveniently be expressed using matrix notations as:

$$\mathbf{f} = \mathbf{T}\hat{\mathbf{f}} \quad \text{with} \quad T_{lk} = \frac{1}{\sqrt{N}} e^{2\pi i k x_l}, \quad (29)$$

where \mathbf{T} is the NDFT matrix. Note that in the case of equispaced spatial nodes such that $x_l = \frac{l}{N}$, this operation corresponds to the conventional DFT and \mathbf{T} reduces to the Fourier matrix \mathbf{F} defined as:

$$F_{lk} = \frac{1}{\sqrt{N}} e^{2\pi i k l / N}. \quad (30)$$

This Fourier matrix is unitary and its inverse is simply its Hermitian conjugate: $\mathbf{F}^{-1} = \mathbf{F}^*$. On the contrary, for non-equispaced spatial nodes \mathbf{x} , the operator \mathbf{T} is typically neither orthogonal nor admits an inverse. Still, one can consider the adjoint NDFT operator \mathbf{T}^* :

$$T_{kl}^* = \frac{1}{\sqrt{N}} e^{-2\pi i k x_l / N}. \quad (31)$$

Although this adjoint operation no longer corresponds to the inverse of the transform, it can be used in practice to estimate the inverse when the problem remains over-determined through a least squares optimisation of the form:

$$\hat{\mathbf{f}} = \arg \min_{\hat{\mathbf{x}}} \|\mathbf{f} - \mathbf{T}\hat{\mathbf{x}}\|_2^2. \quad (32)$$

This problem can be efficiently solved using iterative algorithms (in particular using a conjugate gradient) which involve the computation of both \mathbf{T} and \mathbf{T}^* .

It is therefore important to have fast algorithms for the computation of the NDFT and its adjoint. Note that a naive evaluation of the sum in Eq. (28) would scale as $O(N \times M)$ which is prohibitively large for most applications (including for our mass mapping problem). In this work, we use a fast approximate algorithm¹ to evaluate the NDFT and its adjoint, called NFFT (Keiner et al. 2009), which only scales as $O(N \log(N) + |\log(\epsilon)|M)$ where ϵ is the desired accuracy. For a given sampling of the frequency space (i.e. for a given N), the NFFT only linearly scales with the number M of spatial nodes x_l , which in our case will correspond to the number of galaxies in the survey.

Although the DFT can still be estimated for irregularly sampled data by solving Eq. (32) when the problem remains over-determined, the under-determined problem is ill-posed and cannot be directly solved using a least-squares but requires an additional regularisation.

3.1.2. Sparse regularisation of inverse problems

For the specific application of weak lensing mass mapping, the inverse problem of estimating the Fourier transform of the signal for irregular samples can be severely ill-posed when the galaxy density is low. In that case, a simple least-squares such as Eq. (32) is not able to recover a satisfying estimate of $\hat{\mathbf{f}}$ without additional prior. To regularise this inversion, we propose in this paper to use a so-called analysis-based sparsity prior. We introduce here the concepts of sparse regularisation in a general setting before specialising them to the weak lensing mass mapping problem in the next section.

Let us consider a general linear inverse problem of the form:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (33)$$

where \mathbf{A} is a linear operator, \mathbf{y} are the measurements, contaminated by an additive Gaussian noise \mathbf{n} , and \mathbf{x} is the unknown signal we wish to recover from the measurements \mathbf{y} . Ill-posed inverse problems typically occur when the operator \mathbf{A} is not invertible or extremely badly conditioned, in which case the solution may not be unique and is at least extremely unstable to measurement noise. In order to recover a meaningful estimate of the unknown signal some additional prior information is required to help restrict the space of possible solutions. One particularly

powerful instance of such a prior is the sparsity prior which assumes that the signal can be represented in an suitable domain using only a small number of nonzero coefficients (Starck et al. 2015).

More formally, let us consider a signal \mathbf{x} , it can be analysed using a dictionary Φ to yield a set of analysis coefficients α , such that $\alpha = \Phi^* \mathbf{x}$. For instance, in the case of the Fourier dictionary, Φ would correspond to the DFT matrix \mathbf{F} and α would be the Fourier coefficients of the signal \mathbf{x} . A signal is considered sparse in this analysis framework if only a small number of its coefficients α are non-zero.

Under a sparsity prior, the solution of the inverse problem stated in Eq. (33) can be recovered as the sparsest possible signal $\hat{\mathbf{x}}$ still compatible with the observations. This can be formalised as a convex optimisation problem of the form:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\Phi^* \mathbf{x}\|_1. \quad (34)$$

The first term in this expression in this minimisation problem is a quadratic data fidelity term while the second term promotes the sparsity of the solution by penalising the ℓ_1 norm of the analysis coefficients of \mathbf{x} .

Thanks to very recent advances in the field of convex optimisation and proximal theory, we now have at our disposal very efficient algorithms for solving this optimisation problem, which had remained for a long time largely intractable. In the application presented in this paper, we chose to implement an adaptation of the primal-dual algorithms introduced in Condat (2013), Vu (2013). We direct the interested reader to Appendix A for a few elements of proximal calculus and details concerning this specific algorithm. A more in depth introduction to proximal theory can be found in Starck et al. (2015).

3.2. Sparse recovery algorithm for weak lensing mass mapping

We now consider the specific case of weak lensing mass mapping from irregularly sampled data and we propose a sparse recovery algorithm to reconstruct the convergence map from noisy shear data.

Consider a lensing survey with N_g galaxies. We can write the expression of the shear $\gamma = (\gamma_i)_{i \in \llbracket 0, N_g \rrbracket}$ at the position of each galaxy given a convergence map κ as

$$\gamma = \mathbf{TPF}^* \kappa. \quad (35)$$

In this expression, \mathbf{F} is the Fourier matrix and \mathbf{T} is the NDFT matrix defined for arbitrary spatial nodes \mathbf{x} placed at the position of each galaxy in the survey. The diagonal operator \mathbf{P} implements the transformation from convergence to shear in Fourier space:

$$\hat{\gamma} = \mathbf{P}\hat{\kappa} = \left(\frac{k_1^2 - k_2^2}{k^2} + i \frac{2k_1 k_2}{k^2} \right) \hat{\kappa}. \quad (36)$$

Of course, this expression is not defined for $k_1 = k_2 = 0$, which corresponds to the well known mass-sheet degeneracy and by convention we will set the mean to 0. We are using complex notations for both shear and convergence, with in particular $\kappa = \kappa_E + i\kappa_B$ where κ_E and κ_B are respectively E- and B-modes maps.

This expression is only formally correct if the Fourier transform is evaluated on the \mathbb{R}^2 plane. In practice however, we consider finite fields and rely on a DFT to compute the Fourier transform which imposes artificial periodic conditions. The result of

¹ The C++ library NFFT 3 is available at <https://www-user.tu-chemnitz.de/~potts/nfft>

Eq. (36) is therefore a circular convolution of the convergence field and the Kaiser-Squires kernel. In practice, when evaluating this expression, we apply sufficient zero-padding to the convergence field κ in order to minimise the impact of these periodic boundary conditions. As the Kaiser-Squires kernel decays to the inverse angular distance squared, the quality of the approximation due to the circular convolution quickly improves with the amount of zero-padding.

An important point to stress is that the operator \mathbf{P} is unitary, with $\mathbf{P}^* \mathbf{P} = \text{Id}$, just like the Fourier matrix \mathbf{F} , and as such is readily invertible. Solving Eq. (35) therefore reduces to the inversion of the NDFT operator, which is the only difficult step. As mentioned in the previous section, if the spatial nodes are not regularly spaced, estimating the inverse NDFT is in the general case an ill-posed inverse problem.

Our aim is to apply the sparse regularisation framework introduced in Sect. 3.1.2 to the inversion of the NDFT operator, thus yielding an estimate of the convergence. We consider the following sparse optimisation problem:

$$\bar{\kappa} = \arg \min_{\kappa} \frac{1}{2} \|\Sigma_{\gamma}^{-\frac{1}{2}} [\gamma - \mathbf{TPF}^* \kappa]\|_2^2 + \lambda \|\mathbf{w} \circ \Phi^* \kappa\|_1 + i_{\mathfrak{S}(\cdot)=0}(\kappa). \quad (37)$$

The first term is a quadratic data fidelity term, where Σ_{γ} is the covariance matrix of the individual shear measurements, generally assumed to be diagonal as we are treating galaxies independently. The second term is an analysis-based sparsity constraint where Φ is a dictionary providing a sparse representation of the signal we want to recover, \circ is the Hadamard product (term by term product), and \mathbf{w} is a vector of weights allowing us to adaptively adjust the ℓ_1 ball based on the local level of noise (see Sect. 3.4). Finally, the last term imposes the imaginary part of the solution (i.e. B-modes) to vanish. Remember that the indicator function i_C of a set C is defined by

$$i_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases} \quad (38)$$

Here, $C = \{\kappa \in \mathbb{C}^{N \times N} | \Im(\kappa) = 0\}$ where \Im is the imaginary part, and $N \times N$ is the size of the reconstruction grid. As a result any solution with a non-zero imaginary part is excluded. This additional constraint is crucial as the irregular galaxy sampling leads to important leakage between E- and B-modes. As E-modes are constrained by the sparsity prior, most of the signal would tend to leak towards B-modes without this additional term. Vanishing B-modes is of course a completely physically motivated prior as gravitational lensing only produces E-modes, but this also means that this method will be sensitive to spurious B-modes resulting from uncorrected systematics which will contaminate the recovered signal.

We efficiently solve this problem by adopting a primal-dual algorithm following Condut (2013), Vu (2013). The specialisation of this algorithm to the weak lensing mass mapping problem is presented in Appendix A and summarised in algorithm 1.

The solution of this optimisation problem tends to be biased as a well-known side effect of the ℓ_1 sparsity constraint. In order to correct for this bias and improve the quality of the solution, we implement a two-steps approach.

We first apply the reweighted- ℓ_1 strategy from Candès et al. (2008) which relies on iteratively solving Eq. (37),

Algorithm 1: Analysis-based κ sparse recovery from shear

Require:

Shear of each galaxy in the survey γ .
 Sparsity constraint parameter $\lambda > 0$.
 Weights $w_i > 0$.
 $\tau = 2/(\|\Phi\|^2 + \|\Sigma^{-\frac{1}{2}} \mathbf{T}\|^2)$.

```

1:  $\kappa^{(0)} = 0$ 
2:  $\forall i, \lambda'_i = \lambda w_i$ 
3: for  $n = 0$  to  $N_{\max} - 1$  do
4:    $\nabla^{(n)} = \mathbf{FP}^* \mathbf{T}^* \Sigma_{\gamma}^{-1} (\gamma - \mathbf{TPF}^* \kappa^{(n)})$ 
5:    $\kappa^{(n+1)} = \Re [\kappa^{(n)} + \tau (\nabla^{(n)} - \Phi \alpha^{(n)})]$ 
6:    $\alpha^{(n+1)} = (\text{Id} - \text{ST}_{\lambda'}) (\alpha^{(n)} + \Phi^* (2\kappa^{(n+1)} - \kappa^{(n)}))$ 
7: end for
8: return  $\kappa^{(N_{\max})}$ .
```

adjusting each time the weights \mathbf{w} based on the previous estimate of the solution. This procedure is described below:

1. Set the iteration count $\ell = 0$ and initialise the weights $\mathbf{w}^{(0)}$ according to the procedure described in Sect. 3.4.
2. Solve the weighted ℓ_1 minimisation problem of Eq. (37) using algorithm 1, yielding a solution $\kappa^{(\ell)}$.
3. Update the weights based on the wavelet transform of the solution $\alpha^{(\ell)} = \Phi^* \kappa^{(\ell)}$:

$$w_i^{(\ell+1)} = \begin{cases} \frac{w_i^{(0)}}{|\alpha_i^{(\ell)}|/\lambda w_i^{(0)}} & \text{if } |\alpha_i^{(\ell)}| \geq \lambda w_i^{(0)} \\ w_i^{(0)} & \text{if } |\alpha_i^{(\ell)}| < \lambda w_i^{(0)} \end{cases}, \quad (39)$$

4. Terminate on convergence. Otherwise, increment ℓ and go to step 2.

In practice we find that 3 or 5 re-weightings are generally sufficient to reach a satisfying solution. However, despite improving the quality of the solution, this reweighting scheme does not completely correct for the ℓ_1 amplitude bias.

The second correction step involves using this reweighted- ℓ_1 solution to define the support Ω of the solution in the analysis dictionary, which can then be used in place of the ℓ_1 constraint to recover an unbiased solution. This support is defined as:

$$\Omega_i = \begin{cases} 1, & \text{if } [\Phi^* \kappa^{(\ell)}]_i > \lambda w_i^{(0)} \\ 0, & \text{otherwise.} \end{cases} \quad (40)$$

Using the support to constrain the solution, the recovery problem becomes:

$$\bar{\kappa} = \arg \min_{\kappa} \frac{1}{2} \|\Sigma_{\gamma}^{-\frac{1}{2}} [\gamma - \mathbf{TPF}^* \kappa]\|_2^2 + i_{\Omega}(\Phi^* \kappa) + i_{\mathfrak{S}(\cdot)=0}(\kappa). \quad (41)$$

where i_{Ω} is the indicator function of the support Ω . This problem preserves the sparsity of the solution of the reweighted- ℓ_1 iterative scheme but no longer penalises the amplitude of the analysis coefficients, which leads to an unbiased solution. This problem can still be solved using algorithm 1 simply by replacing the Soft-Thresholding operation on line 6 by a multiplication with the support Ω .

3.3. Choice of dictionary

For any sparse regularisation method, an appropriate choice of dictionary is important to the quality of the result. This is especially true for noise dominated problems where the prior takes

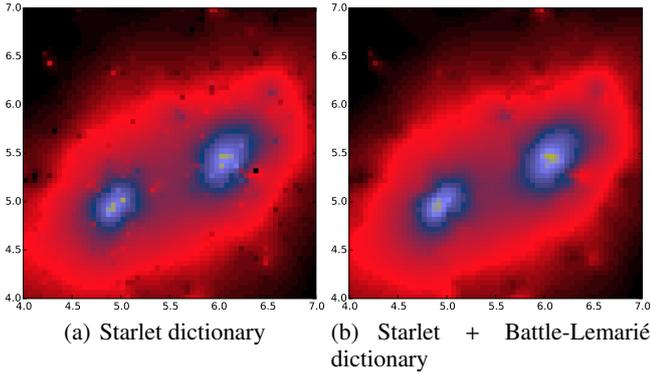


Fig. 3. Comparison of reconstruction using starlets or a combination of starlets and Battle-Lemarié wavelets as the dictionary Φ . The Battle-Lemarié atoms are very effective at penalising isolated pixels.

prevalence when the data is not constraining. Previous sparsity based methods developed for weak lensing mass mapping either employed starlets for denoising (Starck et al. 2006) or DCT for inpainting (Pires et al. 2009). Indeed, at small scale, the non-Gaussian convergence signal essentially generated by isolated galaxy clusters is well represented using the starlet dictionary which features isotropic atoms, adapted to the average circular profile of dark matter halos. On the other hand, on large scales, the DCT is more efficient at capturing the Gaussian part of the convergence signal and has proven to be an excellent dictionary to inpaint missing data due to masks without altering the power spectrum of the reconstructed maps.

In this work, we aim at reconstructing the convergence map on small scales and we therefore adopt the starlet dictionary for the reasons stated above. However, we find that when using the starlet alone, details at the finest scale are not sufficiently constrained, in particular spurious isolated pixels tend to contaminate the solution, as shown on Fig. 3a. To help penalise this unwanted behaviour, we build a hybrid dictionary by concatenating to the starlet dictionary the first scale of an undecimated bi-orthogonal wavelet transform, more specifically a Battle-Lemarié wavelet of order 5. Contrary to starlet atoms, Battle-Lemarié wavelets are much more oscillatory and have a larger support (formally infinite but with an exponential decay). This makes them relatively inefficient at sparsely representing singularities such as isolated pixels, which are therefore more strongly penalised by the sparsity prior.

To illustrate the benefits of using this hybrid dictionary, we compare in Fig. 3 the results of our reconstruction algorithm in a simple noiseless case when using starlets alone or the hybrid dictionary described above, all other parameters being kept fixed. As this simple qualitative comparison demonstrates, starlets alone tend to create small pixel-sized artefacts, even in the absence of noise. These are completely eliminated by the inclusion of the Battle-Lemarié wavelets.

We add that although we find this dictionary to be effective at regularising the mass mapping inversion, it remains generic and was not specifically designed for an optimal representation of convergence maps. More specific dictionaries could be used just as well and would potentially improve further our results. It has for instance been shown that application specific dictionaries built using Dictionary Learning (DL) perform better than wavelets for the recovery of astronomical images Beckouche 2013.

3.4. Adjusting the sparsity constraint

A recurring issue with sparse recovery problems such as the one stated in Eq. (37) is the choice of the regularisation parameter λ . There is unfortunately no general rule indicating how to set this parameter in practice. For this application, we adopt the approach that was proposed in Paykari et al. (2014), which consists in defining this parameter with respect to the noise level.

Formally, the parameter λ scales the ℓ_1 ball used in the sparsity constraint. In practice, it defines the level of Soft Thresholding applied to the dual variable α (see line 6 in algorithm 1) and therefore discriminates between significant and non-significant coefficients. While this threshold can be set according to a given sparsity model of the signal to recover, for noise dominated problems, it is much more crucial to define this threshold with respect to the noise level. As the noise statistics vary across the field, depending on the specific galaxy distribution, we introduce a vector of weights w (found in the ℓ_1 term in Eq. (37)) with the purpose of locally scaling the sparsity constraint based on the standard deviation of the noise propagated to the coefficients α . For each wavelet coefficient α_i we set the weight w_i to the estimated standard deviation $\sigma(\alpha_i)$. As a result, the level of Soft Thresholding applied to each coefficient α_i is $\lambda'_i = \lambda w_i$ and accounts for noise variations across the field. As the threshold is proportional to the standard deviation of the noise, it can be interpreted as an hypothesis test to determine if a coefficient is due to signal or noise, assuming Gaussian statistics for the noise, which adds a powerful detection aspect to the sparsity constraint.

To estimate this noise level, and therefore set the weights w , it is first necessary to understand how the noise in the data propagates to the dual variable α . By considering algorithm 1, it can be seen that the noise at the level of the shear γ_N is propagated to the wavelet coefficients through the operation $\Phi^* \mathcal{R} (\mathbf{F} \mathbf{P}^* \mathbf{T}^* \gamma_N)$. In practice, we estimate the standard deviation of wavelet coefficients by generating Monte-Carlo noise simulations, obtained by keeping the galaxies at their observed position while randomising their orientation. Note that this step needs only to be performed once, outside of the main iteration of algorithm 1. An example of the resulting standard deviation maps for different wavelet scales is shown on Fig. 4. As can be seen, on the finest scales, the noise level has important fluctuations across the field and the contribution of individual galaxies can be seen. On larger scales, these local fluctuations are smoothed out and the noise level becomes much more homogeneous. Using this strategy therefore allows us to tune locally the sparsity constraint to take into account the specific galaxy distribution of the survey and leaves only one free parameter λ .

3.5. Numerical experiment on noiseless data

The algorithm presented in this section is only meant to address the irregular sampling of the shear and the presence of noise, the complete problem being solved in the next section. In this simplified setting, we present a small numerical experiment to verify the algorithm's effectiveness at solving the linear inverse problem.

We simulated a 10×10 arcmin² field, containing a group of galaxy clusters extracted from the Bolshoi N -body simulations (Klypin et al. 2011, see Sect. 5.1, for more details), as shown on Fig. 5. These clusters are placed at redshift $z_1 = 0.3$ and we simulate a lensing catalogue with randomly distributed sources on a single lens plane at redshift $z_s = 1.2$. Note that we only simulated shear measurements from the input convergence map and not the reduced shear.

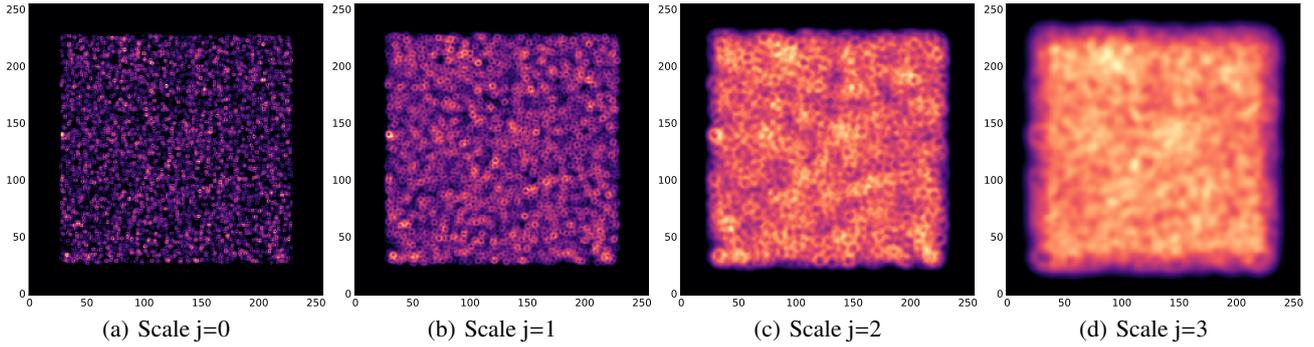


Fig. 4. Standard deviation maps w used to scale the sparsity constraint, obtained by propagating the shear noise to the wavelet coefficients. We show here noise maps for four successive starlet scales.

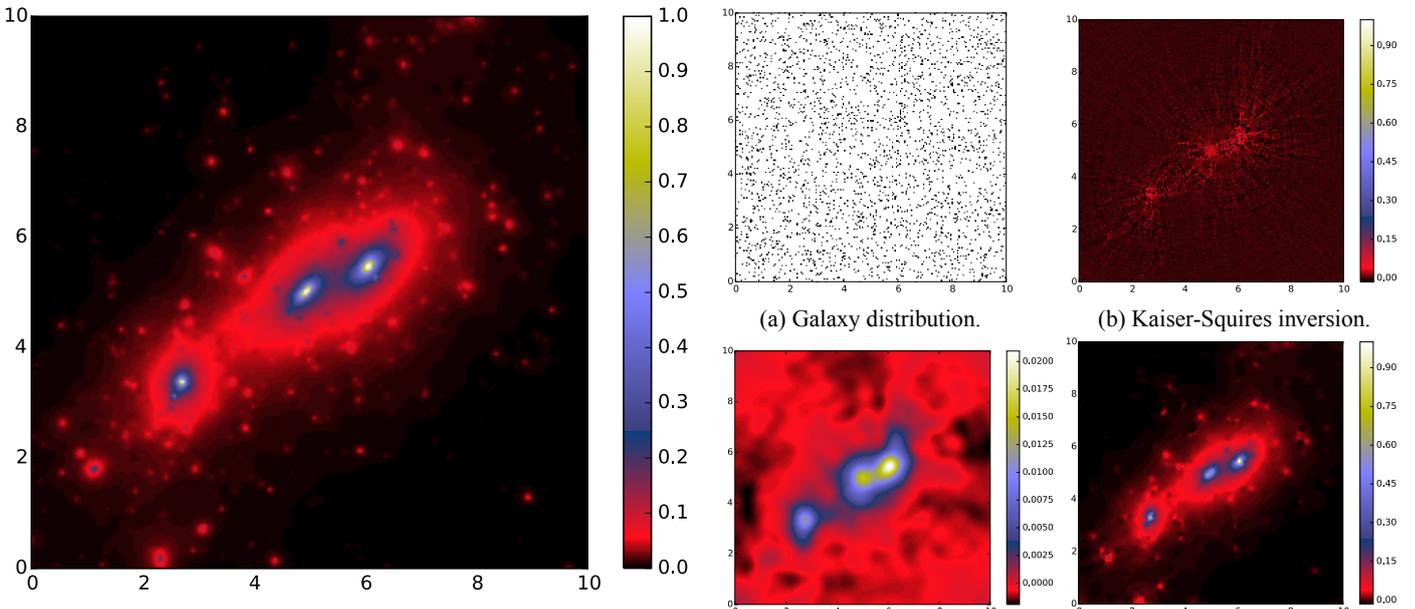


Fig. 5. Test convergence map generated from an N -body simulation. The clusters are located at $z_1 = 0.3$ while the source plane is placed at $z_s = 1.2$.

We consider here the noiseless inpainting problem where the input shear is exactly known at the position of the sources. Of course this problem is unrealistic but does illustrate the impact of missing data on the mass mapping inversion using Fourier estimators.

We generated a shear catalogue with a density of 30 galaxies per square arcminute which corresponds to a typical Euclid density. To exacerbate the effects of missing data, we reconstructed the input convergence map on a grid with 3 arcsec pixels. Binning the input catalogue at this resolution corresponds to an average of 0.075 galaxies per pixels, which means mostly empty pixels as illustrated by the mask in Fig. 6a where empty pixels are shown in white. A direct Kaiser-Squires inversion with such a mask is shown on Fig. 6b which corresponds to the solution of the inverse problem in the absence of regularisation. In order to recover the main structures we also show the result of a Gaussian smoothing with a kernel of 0.25 arcmin on Fig. 6c. Note that on this last figure the colour scale is adjusted for better visualisation but the amplitude of the reconstruction is an order of magnitude below the input signal smoothed with the same kernel. As illustrated by these plots, small scale information is lost

(c) Kaiser-Squires with a 0.25 arcmin Gaussian smoothing.

(d) Sparse recovery.

Fig. 6. Reconstruction of the input convergence map using 30 galaxies per square arcminute using a Kaiser-Squires estimator **b–c**) and sparse recovery **d**). The *top left panel* shows the mask applied to the binned shear map with empty pixels marked in white. In this test, 93% of the pixels are masked.

with this approach and although using larger bins would regularise the inversion these details would still be lost.

We then applied [algorithm 1](#) to the same catalogue. As the weighting scheme presented in [Sect. 3.4](#) is not meant to be used on noiseless data, we use a noisy version of the shear data to estimate the weights w and we set the regularisation parameter λ to a very low value ($\lambda = 0.01$) to apply the algorithm on noiseless data. The solution of the optimisation problem is shown on [Fig. 6](#). The sparse recovery algorithm constitutes a major improvement over a simple Kaiser-Squires inversion and is able to recover pixel scale details despite 93% of the field being masked. Several reasons can explain these surprising results. First and foremost, the shear information is non local and even small structures formally impact the shear across all the field, albeit with an amplitude decaying to the square of the angular distance. The second reason is the use of *isotropic* wavelets which provide a sparse representation of the true convergence

map while being morphologically distinct from the response of individual galaxies, which is quadrupolar. As a result, the sparsity constraint greatly favours the true convergence map, which is smooth with small isotropic features, over anisotropic artefacts resulting from the irregular shear sampling.

The point of this numerical experiment is to show that although the shear field is randomly sampled at a relatively low rate, high frequency information can still be recovered using our method. Of course, in practice small-scale details are generally lost in the considerable amount of noise coming from intrinsic galaxy ellipticities, but not necessarily at the close vicinity of the center of galaxy clusters where the shear signal can become significant even on small scales. The ability to reconstruct small-scale details is even more relevant when including flexion information as the noise power spectrum of the flexion estimator drops at small scales.

4. Sparse recovery for the complete problem: the Glimpse2D algorithm

In the previous section, we have introduced an algorithm based on sparse regularisation for solving the simplified linear inversion problem. Although not realistic, this problem is still an important step towards the complete surface density reconstruction that we address now in this section. We detail how the algorithm of the previous section can be modified to take into account reduced shear, redshift information for individual galaxies, and flexion information.

4.1. Handling the reduced shear

The problem addressed in the previous section assumed knowledge of the shear, in which case the inverse problem remains linear. While this assumption can be made in the weak regime, it breaks down at the vicinity of the structures (galaxy clusters) we are interested in mapping. Although the method presented in the previous section no longer directly applies, we present in this section how it can be extended to take into account the reduced shear $g = \frac{\gamma}{1-\kappa}$, which makes the inversion problem non-linear.

Throughout this paper, we will restrict ourselves to the case $|g| \leq 1$ for simplicity, assuming that the sources which do not verify this condition can be identified and excluded from the sample. The method presented here could be extended using an iterative procedure to identify and thus treat accordingly sources lying in the region $|g| > 1$.

Replacing the shear by the reduced shear in the inversion problem stated in Eq. (37) yields:

$$\arg \min_{\kappa} \frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} \left[g - \frac{\mathbf{TPF}^* \kappa}{1 - \mathbf{TF}^* \kappa} \right] \right\|_2^2 + \lambda \|\mathbf{w} \circ \Phi^t \kappa\|_1 + i_{\mathfrak{S}(\cdot)=0}(\kappa). \quad (42)$$

Note that at the denominator the NDFT operator is only used to evaluate the convergence at the position of each galaxy. In this form, the full problem cannot be directly addressed using the algorithm presented in the previous section as the operator to invert is no longer linear. Nonetheless, following a common strategy to handle this non-linearity, the term $C_{\kappa}^{-1} = \Sigma^{-\frac{1}{2}} / (1 - \mathbf{TF}^* \kappa)$ can be factored out and be interpreted as a diagonal covariance matrix, which depends on the signal κ :

$$\arg \min_{\kappa} \frac{1}{2} \left\| C_{\kappa}^{-1} [(1 - \mathbf{TF}^* \kappa)g - \mathbf{TPF}^* \kappa] \right\|_2^2 + \lambda \|\mathbf{w} \circ \Phi^t \kappa\|_1 + i_{\mathfrak{S}(\cdot)=0}(\kappa). \quad (43)$$

If the factor C_{κ}^{-1} is kept fixed, the problem is now linear and can be solved once again using the same class of algorithms as in the previous section. By iteratively solving this linearised problem, updating each time the matrix C_{κ}^{-1} with the current estimate of κ we can recover the solution of the original problem stated in Eq. (42). This is for instance the strategy adopted in Merten et al. (2009).

4.2. Including redshift information

So far we have not given any consideration to the fact that lensing sources are not located on a single plane but are distributed in redshift. As was described in Sect. 2.4, for a lens at a given redshift, the amplitude of the lensing effect experienced by each source will depend on its own redshift. Let us note \mathbf{Z} the diagonal matrix of weights Z_i introduced in Sect. 2.4, the reduced shear can be computed from the convergence κ using the Fourier operators introduced thus far as:

$$g = \frac{\mathbf{ZTPF}^* \kappa}{1 - \mathbf{ZTF}^* \kappa}, \quad (44)$$

where κ is understood to be the convergence at infinite redshift. With this new operator, the full inversion problem becomes:

$$\arg \min_{\kappa} \frac{1}{2} \left\| C_{\kappa}^{-1} [(1 - \mathbf{ZTF}^* \kappa)g - \mathbf{ZTPF}^* \kappa] \right\|_2^2 + \lambda \|\mathbf{w} \circ \Phi^t \kappa\|_1 + i_{\mathfrak{S}(\cdot)=0}(\kappa). \quad (45)$$

where the matrix C_{κ}^{-1} now becomes $C_{\kappa}^{-1} = \Sigma^{-\frac{1}{2}} / (1 - \mathbf{ZTF}^* \kappa)$. This is simply a generalisation of Eq. (43), which can be recovered when no redshift information is available by setting $\mathbf{Z} = \text{Id}$, and can be solved exactly in the same way.

The main advantage of using redshift estimates for individual galaxies is the proper scaling of the resulting convergence map, which can be translated into a physical surface mass density map Σ of the lens plane:

$$\Sigma(\theta) = \kappa(\theta) \Sigma_{\text{crit}}^{\infty}. \quad (46)$$

The mass of the lens can then be estimated by integrating Σ within a given radius. The second advantage of using individual redshifts is that it can help mitigate the mass-sheet degeneracy, as was described in Bradac et al. (2004) and presented in Sect. 2.4. We stress however that this degeneracy cannot be completely lifted using the method presented here as, of the two terms in the gradient of the χ^2 in Eq. (A.14), we only retain the one that is insensitive to the mean. Therefore, the mean value of the field remains unconstrained by the data using our algorithm. Nevertheless, we expect the additional redshift information to locally break the degeneracy and help recover the correct amplitude for the most significant structures.

4.3. Improving angular resolution with flexion

We demonstrated in the previous section that our sparse recovery algorithm is capable of reconstructing small scale details in the absence of noise despite the irregular galaxy sampling. In practice however, the shear is noise dominated on those scales, which makes recovering high frequency information from shear measurements unlikely.

As was explained in Sect. 2.2, while the Kaiser-Squires estimator for shear has a flat noise power spectrum, the noise power spectrum of the minimum variance estimator for flexion has a

$1/k^2$ dependency. As a result, although flexion measurements are generally very noisy, the noise level of the estimated map eventually drops below that of the shear on sufficiently small scales. Flexion can therefore bring useful information but only below a given scale and is very complementary to shear.

Our aim is therefore to improve the mass-map reconstruction on small scales by extending our reconstruction method to incorporate flexion information. In the interest of simplicity, we consider here only the linear problem without redshift information, to highlight the difficulties inherent to the inclusion of flexion. The full problem is solved in the next section.

Following the approach developed in the first section, we address the problem using Fourier estimators. We first introduce the diagonal operator \mathbf{Q} implementing the transform from convergence κ to first flexion \mathcal{F} in Fourier space, defined as:

$$\hat{\mathcal{F}} = \mathbf{Q}\hat{\kappa} = (k_2 - ik_1)\hat{\kappa}, \quad (47)$$

where we use complex notations for the flexion with $\mathcal{F} = \mathcal{F}_1 + i\mathcal{F}_2$. Contrary to the shear operator \mathbf{P} , this operator \mathbf{Q} is not unitary but is still invertible:

$$\hat{\kappa} = \mathbf{Q}^{-1}\hat{\mathcal{F}} = \frac{\mathbf{Q}^*}{k^2}\hat{\mathcal{F}} = \frac{k_2 + ik_1}{k^2}\hat{\mathcal{F}}. \quad (48)$$

To extend the algorithm presented in the previous section, a first straightforward approach would be to simply add a flexion term to the χ^2 of Eq. (37) and solve the following problem:

$$\arg \min_{\kappa} \frac{1}{2} \|\gamma - \mathbf{TPF}^*\kappa\|_2^2 + \frac{1}{2} \|\mathcal{F} - \mathbf{TQF}^*\kappa\|_2^2 + \lambda \|\mathbf{w} \circ \Phi^t\kappa\|_1 + i_{\mathfrak{S}(\cdot)=0}(\kappa). \quad (49)$$

Although formally correct, solving this problem using the primal-dual algorithm presented in the previous section leads to a number of technical issues linked to the fact that the operator \mathbf{Q} , being not unitary, now contributes to the difficulty of the inverse problem. In particular, this impacts our ability to robustly identify significant coefficients in the gradient of the data fidelity term (key to the regularisation strategy presented in the previous section), which is now affected by a mixture of a flat shear noise and a flexion noise with a power spectrum in k^2 . Furthermore, even without considering the regularisation, the inversion of the operator \mathbf{Q} , which is essentially a 2D gradient, requires a large number of iterations if solved with a standard gradient descent. These considerations make the algorithm much slower and far less robust to noise than when solving the problem from shear alone.

This needs not be the case however as the inverse of this operator is explicit in Fourier space and these difficulties can be avoided if we make proper use of this explicit inverse. We propose therefore to address the combined shear and flexion reconstruction as the following sparse optimisation problem:

$$\arg \min_{\kappa, \tilde{\mathcal{F}}} \frac{1}{2} \|\gamma - \mathbf{TPF}^*\kappa\|_2^2 + \frac{1}{2} \|\mathcal{F} - \mathbf{TF}^*\tilde{\mathcal{F}}\|_2^2 + \lambda \|\mathbf{w} \circ \Phi^t\kappa\|_1 + i_{\text{Im}(\mathbf{R})} \left(\begin{pmatrix} \kappa \\ \tilde{\mathcal{F}} \end{pmatrix} \right), \quad (50)$$

where we introduce the application $\mathbf{R} : \mathbb{R}^{N \times N} \rightarrow \mathbb{C}^{2N \times N}$, $\kappa \mapsto \begin{pmatrix} \kappa \\ \mathbf{FQF}^*\kappa \end{pmatrix}$, with $N \times N$ the size of the reconstruction grid.

Thanks to the inclusion of the auxiliary variable $\tilde{\mathcal{F}}$ we have now decoupled the problem of the inversion of the NDFT operator \mathbf{T} from the conversion between flexion and convergence which is now addressed implicitly in the last term. Remember that the indicator function i_C is infinite outside of the set C and therefore exclude any solution which do not belong to C . In our case, we require the solution to be in the image of the operator \mathbf{R} :

$$\text{Im}(\mathbf{R}) = \left\{ \begin{pmatrix} \kappa \\ \tilde{\mathcal{F}} \end{pmatrix} \in \mathbb{C}^{2N \times N} \mid \exists \kappa \in \mathbb{R}^{N \times N}, \tilde{\mathcal{F}} = \mathbf{FQF}^*\kappa \right\}. \quad (51)$$

The constraint $i_{\text{Im}(\mathbf{R})}$ therefore implies two conditions. First, the recovered convergence has to be real, which is equivalent to enforcing the vanishing B-modes condition already used for the shear alone inversion problem. The second is that the recovered flexion $\tilde{\mathcal{F}}$ needs to match the flexion derived from the recovered convergence, which makes the connection between the two variables κ and $\tilde{\mathcal{F}}$. Interestingly, the implementation of this constraint in the primal-dual algorithm naturally gives rise to the minimum variance estimator for shear and flexion introduced in Eq. (17) and therefore will ensure that shear and flexion are optimally combined at each iteration of the algorithm to maximise the S/N of features present in the signal on all scales.

4.4. Complete surface density mapping algorithm

We now present the complete reconstruction algorithm, called Glimpse2D, combining reduced shear \mathbf{g} and reduced flexion \mathbf{F} , and taking into account individual redshift estimates for the sources. The complete problem we aim to solve is the following:

$$\arg \min_{\kappa, \tilde{\mathcal{F}}} \frac{1}{2} \|C_{\kappa\mathbf{g}}^{-1} [(1 - \mathbf{ZTF}^*\kappa)\mathbf{g} - \mathbf{ZTPF}^*\kappa]\|_2^2 + \frac{1}{2} \|C_{\kappa\mathbf{F}}^{-1} [(1 - \mathbf{ZTF}^*\kappa)\mathbf{F} - \mathbf{ZTF}^*\tilde{\mathcal{F}}]\|_2^2 + \lambda \|\mathbf{w} \circ \Phi^t\kappa\|_1 + i_{\text{Im}(\mathbf{R})} \left(\begin{pmatrix} \kappa \\ \tilde{\mathcal{F}} \end{pmatrix} \right), \quad (52)$$

where the non-linear correction matrices $C_{\kappa\mathbf{g}}^{-1}$ and $C_{\kappa\mathbf{F}}^{-1}$ incorporate the diagonal covariance matrices of the shear and flexion measurement respectively. Note that by separating shear and flexion terms, we are assuming that these components are independent. We use this simplistic assumption as the amplitude of the correlation between shear and flexion measurements has not been yet quantified in practice for the AIM method. However, our formalism is trivially extensible to the full non diagonal covariance between shear and flexion, should it be provided.

As in the previous section, the problem can be solved using the same primal-dual algorithm derived from [Condat \(2013\)](#), [Vu \(2013\)](#). The specialisation of this algorithm to the full mass mapping problem of Eq. (52) is provided in [algorithm 2](#) and derived in [Sect. A.3](#).

Just as with the linear problem, we apply a reweighted- ℓ_1 strategy to correct for the bias caused by the ℓ_1 sparsity constraint. As the non-linear correction also requires to iteratively solve this problem we combine the update of the weights \mathbf{w} and the update of the matrix C_κ in the following iterative procedure:

1. Set the iteration count $\ell = 0$, $C_\kappa^{(0)} = 1.0$ and initialise the weights $\mathbf{w}^{(0)}$ according to the procedure described in [Sect. 3.4](#).
2. Solve the weighted ℓ_1 minimisation problem of Eq. (52) using [algorithm 2](#), yielding a solution $\kappa^{(\ell)}$.

Algorithm 2: Analysis-based surface density mapping algorithm from reduced shear and flexion

Require:

Reduced shear and flexion of each galaxy in the survey \mathbf{g} and \mathbf{F} .
Redshift weights for each galaxy in the survey \mathbf{Z} .
Reduced shear correction matrix \mathbf{C}_{κ}^{-1} .
Sparsity constraint parameter $\lambda > 0$.
Weights $w_i > 0$.
 $\tau = 2/(\|\Phi\|^2 + \|\Sigma^{-\frac{1}{2}}\mathbf{T}\|^2)$.

1: $\kappa^{(0)} = 0$; $\tilde{\mathcal{F}}^{(0)} = 0$
2: $\forall i, \lambda'_i = \lambda w_i$
3: **for** $n = 0$ to $N_{\max} - 1$ **do**
4: $\nabla^{(n)} = \mathbf{F}\mathbf{P}^*\mathbf{T}^*\mathbf{Z}\mathbf{C}_{\kappa g}^{-2} \left((1 - \mathbf{Z}\mathbf{T}\mathbf{F}^*\kappa^{(n)})\mathbf{g} - \mathbf{Z}\mathbf{T}\mathbf{P}\mathbf{F}^*\kappa^{(n)} \right) + \mathbf{F}\mathbf{T}^*\mathbf{Z}\mathbf{C}_{\kappa F}^{-2} \left((1 - \mathbf{Z}\mathbf{T}\mathbf{F}^*\kappa^{(n)})\mathbf{F} - \mathbf{Z}\mathbf{T}\mathbf{P}\mathbf{F}^*\tilde{\mathcal{F}}^{(n)} \right)$
5: $\begin{pmatrix} \kappa^{(n+1)} \\ \tilde{\mathcal{F}}^{(n+1)} \end{pmatrix} = \text{prox}_{\text{Im}(\mathbf{R})} \left(\begin{pmatrix} \kappa^{(n)} \\ \tilde{\mathcal{F}}^{(n)} \end{pmatrix} + \tau \left(\nabla^{(n)} - \mathbf{R}\Phi\alpha^{(n)} \right) \right)$
6: $\alpha^{(n+1)} = (\text{Id} - \text{ST}_{\lambda'}) \left(\alpha^{(n)} + \Phi^t \left(2\kappa^{(n+1)} - \kappa^{(n)} \right) \right)$
7: **end for**
8: **return** $\kappa^{(N_{\max})}$.

3. Update the matrices $\mathbf{C}_{\kappa g}^{(\ell)} = \Sigma_g^{-1}(1 - \mathbf{Z}\mathbf{T}\mathbf{F}^*\kappa^{(\ell)})$ and $\mathbf{C}_{\kappa F}^{(\ell)} = \Sigma_F^{-1}(1 - \mathbf{Z}\mathbf{T}\mathbf{P}\mathbf{F}^*\kappa^{(\ell)})$.
4. Update the weights based on the wavelet transform of the solution $\alpha^{(\ell)} = \Phi^t \kappa^{(\ell)}$:

$$w_i^{(\ell+1)} = \begin{cases} \frac{w_i^{(0)}}{|\alpha_i^{(\ell)}|/\lambda w_i^{(0)}} & \text{if } |\alpha_i^{(\ell)}| \geq \lambda w_i^{(0)} \\ w_i^{(0)} & \text{if } |\alpha_i^{(\ell)}| < \lambda w_i^{(0)} \end{cases}, \quad (53)$$

5. Terminate on convergence. Otherwise, increment ℓ and go to step 2.

As for the linear problem we find that 3 to 5 iterations of this procedure are generally sufficient to reach a satisfying solution. To complement the reweighting scheme, the same de-biasing step based on the support of the reweighted solution is applied to yield the final solution.

5. Results on simulations

In this section, the performance of the reconstruction algorithm is assessed on lensing simulations using realistic clusters extracted from N -body simulations and realistic noise levels for a space based shear and flexion survey.

5.1. Lensing simulations

In order to assess the performance of the algorithm, we built a test data set based on N -body simulations. Three massive clusters were extracted at $z = 0$ from the Bolshoi simulations (Klypin et al. 2011) using the CosmoSim² interface. These clusters were selected because of their complex geometry with substantial substructure. In each case, at least one halo above $10^{13} h^{-1} M_{\odot}$ can be found within the virial radius of the central halo. The masses of the three clusters we consider in this work can be found in Table 1.

² <http://www.cosmosim.org>

Table 1. Parameters of the three halos extracted from the Bolshoi simulation.

Field number	Virial mass [$h^{-1}M_{\odot}$]	Virial radius [h^{-1} Mpc]
1	1.09×10^{15}	2.14
2	3.02×10^{14}	1.43
3	2.70×10^{14}	1.45

Density maps were obtained for these three clusters by projecting and binning the particle data with a high resolution, corresponding to pixels with an angular size of 0.5 arcsec, for clusters located at redshift $z_L = 0.3$. A multiscale Poisson denoising was subsequently applied to the binned density maps to remove the shot noise from the numerical simulation. The resulting surface mass density maps were then scaled to create convergence maps corresponding to clusters at a redshift $z_L = 0.3$ lensing background galaxies at infinite redshift. Although we acknowledge that these three particular clusters, selected at $z = 0$, are not necessarily representative of clusters at $z_L = 0.3$, for the purpose of testing the mapping algorithm they provide more realistic density distributions than simple models based on SIS or NFW profiles. Shear and flexion maps were then derived from these reference convergence maps and only the center 10×10 arcmin central region of each field is kept. At the redshift of the lenses, this corresponds to a physical size of $1.88 \times 1.88 h^{-1}$ Mpc. Figure 7 illustrates the convergence maps scaled for sources at infinite redshift for the 3 fields we consider.

Finally, 100 mock galaxy catalogues were produced for the three fields using for each realisation a uniform spatial distribution of background galaxies with a density of 80 gal/arcmin² and the following redshift distribution:

$$p(z) \propto \frac{z^2}{2z_0^3} \exp(-z/z_0), \quad (54)$$

with $z_0 = 2/3$. The median redshift of this distribution is $z_{\text{med}} = 1.75$ and we truncate the distribution at $z_{\text{max}} = 5$. This particular distribution has been used in a number of different works and in particular in Cain et al. (2016) to represent the actual galaxy distribution for a typical HST/ACS field. We compute the reduced shear and flexion for each source galaxy based on their redshift and on the redshift of the lens. In the final mock catalogues we assume Gaussian photometric redshifts errors for each sources with $\sigma_z = 0.05(1 + z)$, an intrinsic shape noise of $\sigma_{\epsilon} = 0.3$ for the reduced shear measurements and $\sigma_F = 0.029 \text{ arcsec}^{-1}$ for the reduced flexion measurements. This particular value for the flexion noise is in accordance with previous works (Rowe et al. 2013) and corresponds to the median dispersion of the flexion measurements obtained using the AIM method on HST data for the Abel 1689 cluster (Cain et al. 2011).

Some final cuts are applied to the galaxy catalog to exclude strongly lensed sources by discarding all sources which verify $|g| \geq 1$ or $|F| \geq 1.0 \text{ arcsec}^{-1}$. The flexion cut is based on recommendations from Cain et al. (2016) and constitutes a simple approximation to the practical limit encountered in real data when estimating flexion for extremely lensed sources.

This particular setting was chosen to match the simulations used in Cain et al. (2016) so that our results can be qualitatively compared to theirs. However, we stress that contrary to their work, our reconstruction method relies only on shear and flexion

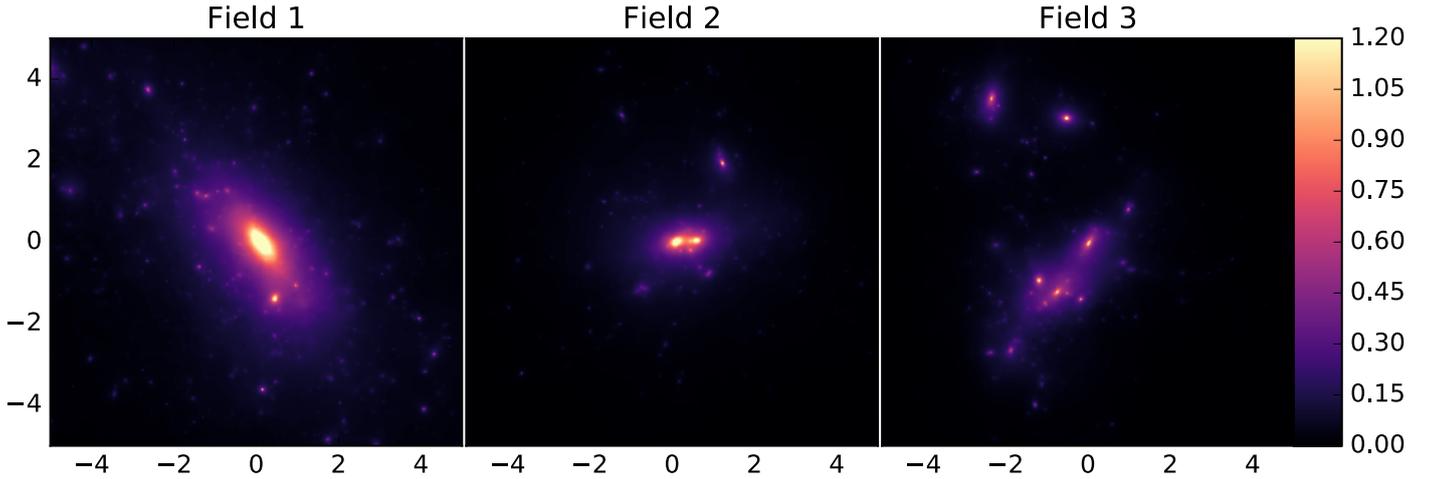


Fig. 7. Convergence maps for the three test clusters extracted from the N -body simulation, assuming a lens redshift of $z_L = 0.3$ and background sources at infinite redshift.

Table 2. Parameters of the reconstruction algorithm.

Parameter	Value
Pixel size	0.05 arcmin
Number of wavelet scales	7
K -sigma threshold	5
Number of iterations	500
Number of re-weightings	5

and does not include strong lensing constraints which are very powerful to map the very center of the clusters.

Finally, throughout this section, we assume a fiducial Λ CDM model for computing distances with $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Cosmological computations are implemented using the NICA EA library³.

5.2. Results

For each of the mock catalogues generated, we performed the inversion with and without the flexion information to assess how much the flexion information can help recover the substructure of the clusters. The same parameters are used for all the three different fields and are summarised in Table 2. The pixel size can in theory be arbitrarily small but we choose 0.05 arcmins as a good compromise between resolution and computational cost. The number of wavelet scales is not crucial to the quality of the reconstruction, we adjust it so that the maximum scale corresponds roughly to the order of the maximum size of the structures we want to recover. Finally, the number of iterations and reweighting steps is not crucial either as the algorithm does converge to a solution.

Assessing the quality of the inversion from a single realisation of the galaxy catalog is difficult as the ability to detect small structure will depend on the specific positions of the galaxies in one realisation. Therefore, we computed the mean and standard deviation of the reconstructed maps over 100 realisations of both shape noise and galaxy positions. The mean of the reconstructions for the 3 fields with and without flexion is represented in Fig. 8. The most striking difference between the 2 sets of reconstructions is the flexion’s ability to detect very small substructures at the 10 arcsec scale which are not

detected from shear alone. This is clearly visible for the first and third fields.

The errors on the reconstructed mass maps are estimated by taking the standard deviation of each pixel in the 100 independent mock catalogues realisation. It is important to stress that the observed dispersion of the reconstructions are due to 2 effects: different noise realisations and different galaxy distributions in angular position and redshift. The standard deviation with and without flexion for all 3 fields is illustrated in Fig. 10. As can be seen on this figure, we only observe significant dispersion on small scales, around detected structures. This is expected as these scales have the lowest signal to noise ratio in the data which makes the reconstruction of these features strongly dependent on the specific noise and galaxy distribution realisation. Nevertheless, we find that the reconstruction is remarkably stable between realisations. Figure 9 shows reconstructions of the mass maps for one realisation of the galaxy catalogues. We see that for a single realisation the reconstructed maps is very close to the mean maps and mainly differs for small substructure. For instance, the sub-halo at the bottom of field 3 appears stronger in this particular realisation than in the average maps. Such deviations are unavoidable but the shape of the halos can reliably be reconstructed from a single data set, as would be the case for actual data.

Flexion does not only enable the detection of very small structures, it also helps to constrain the small scale shape of the main halos. Indeed, in shear reconstructions, the noise dominates the signal on these scales which makes the sparsity prior apparent on small scales. As we are using isotropic wavelets, without sufficient evidence from the data, the reconstruction will be biased towards isotropic shape. This is visible for instance at the center of the first halo, on the top left image of Fig. 8, where the elongation of the very center of the cluster, visible in Fig. 7, is clearly lost. On the contrary, including flexion information helps to constrain the shape of the center of the cluster.

The same effect can be seen in Fig. 11 where we show the absolute difference between the mean of the reconstructions and input convergence maps. The clusters are successfully recovered on large scales in both cases, leaving mostly only the very fine substructure visible in these residual maps. Nevertheless, although small, some residuals of the main halos remain visible near the center of the center fields but are lower when combining shear and flexion, indicating better recovery of the shape of the central part of the cluster when flexion is included. This

³ <http://www.cosmostat.org/software/nicaea>

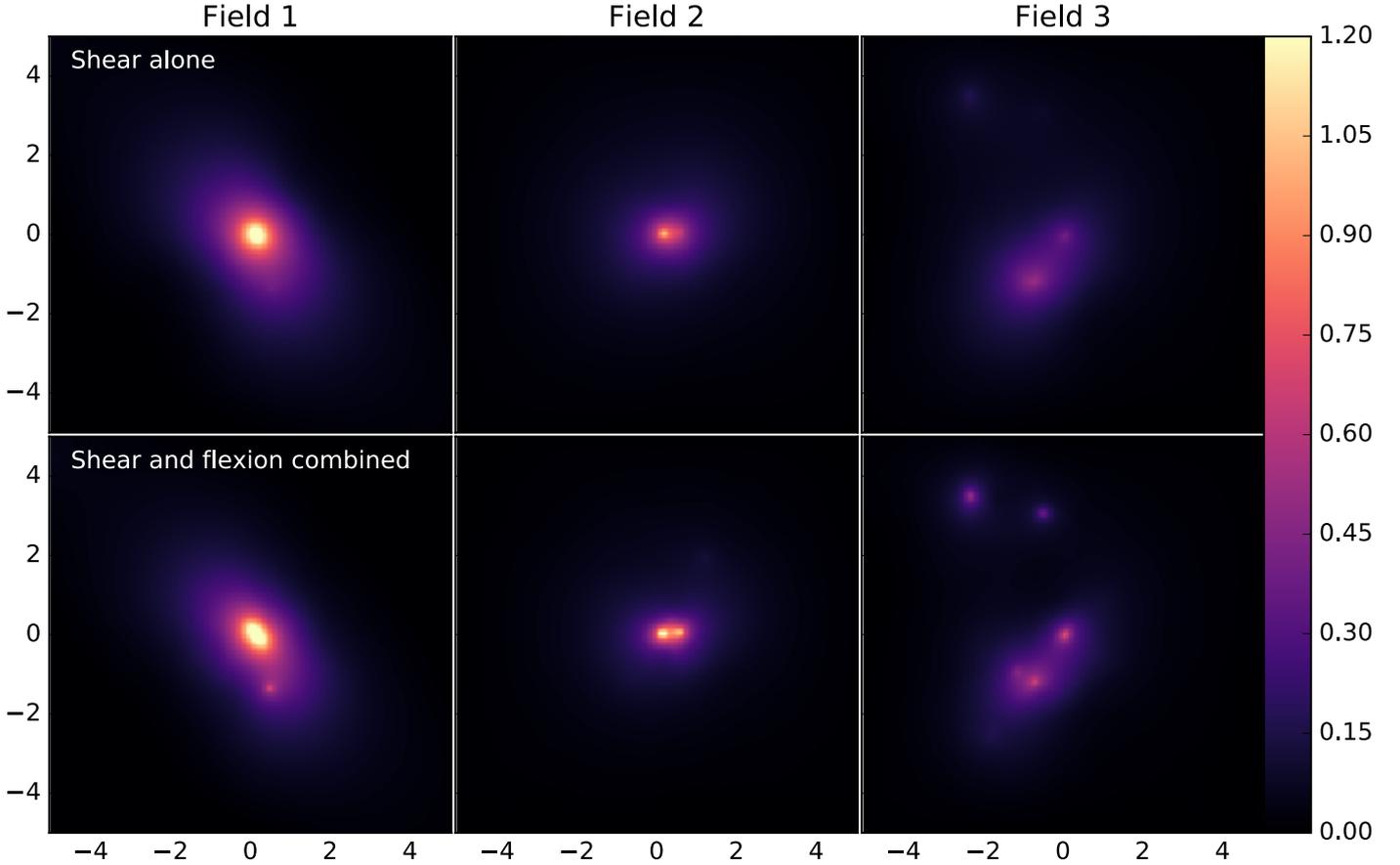


Fig. 8. Mean of 100 independent signal realisations for the three fields. The *top row* corresponds to reconstructions using the shear alone while the *bottom row* corresponds to reconstructions using shear and flexion.

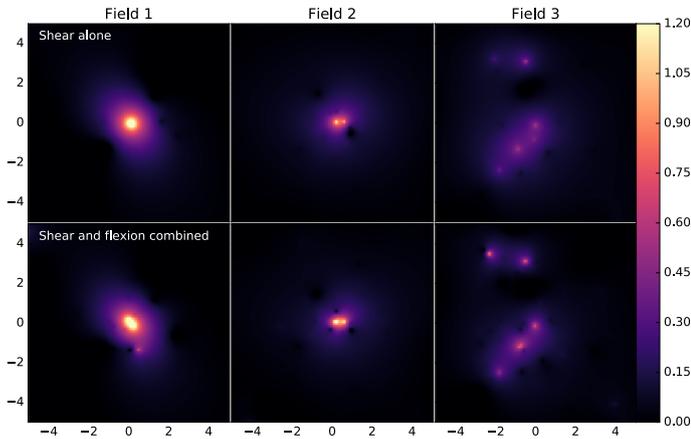


Fig. 9. Reconstruction of the 3 clusters for a single noise realisation. The *top row* corresponds to reconstructions using the shear alone. The *bottom row* corresponds to simulations using shear and flexion.

is particularly visible for the first cluster where the shear only residuals clearly indicate that the anisotropy of the halo is not captured as well by the shear alone reconstruction than by the combined shear and flexion reconstruction.

We also note that on large scales, where the shear dominates, the two sets of reconstructions agree with each other and the inclusion of flexion does not modify the reconstruction. This is illustrated by Table 3 where the integrated mass within a $1'$ radius is computed on the reconstructions with and without flexion. As

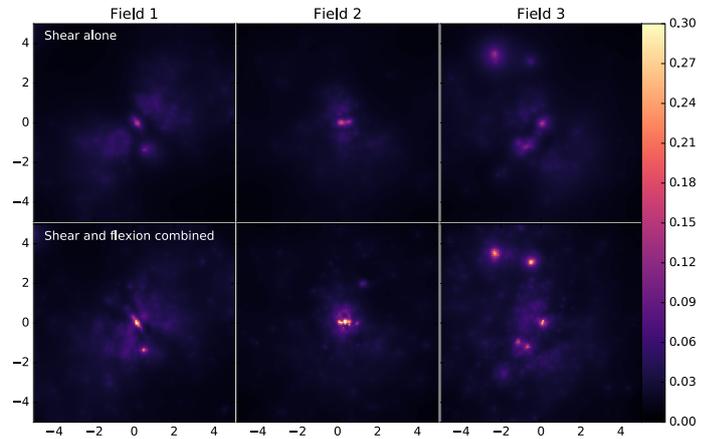


Fig. 10. Standard deviation of the 100 independent signal realisations for the 3 clusters. The *top row* corresponds to reconstructions using the shear alone. The *bottom row* corresponds to simulations using shear and flexion.

can be seen, tight constraints on the halo masses can be obtained for the three fields considered from shear only and the addition of flexion does not change the mass estimates on scales larger than the arcminute.

To quantify the impact of flexion on the recovery of small-scale structures, we show in Fig. 12 aperture masses for five different radii ($19, 38, 75, 150,$ and $300 h^{-1}$ kpc or $6'', 12'', 24'', 48'', 1.6'$) at the center of the 3 fields and compare them with

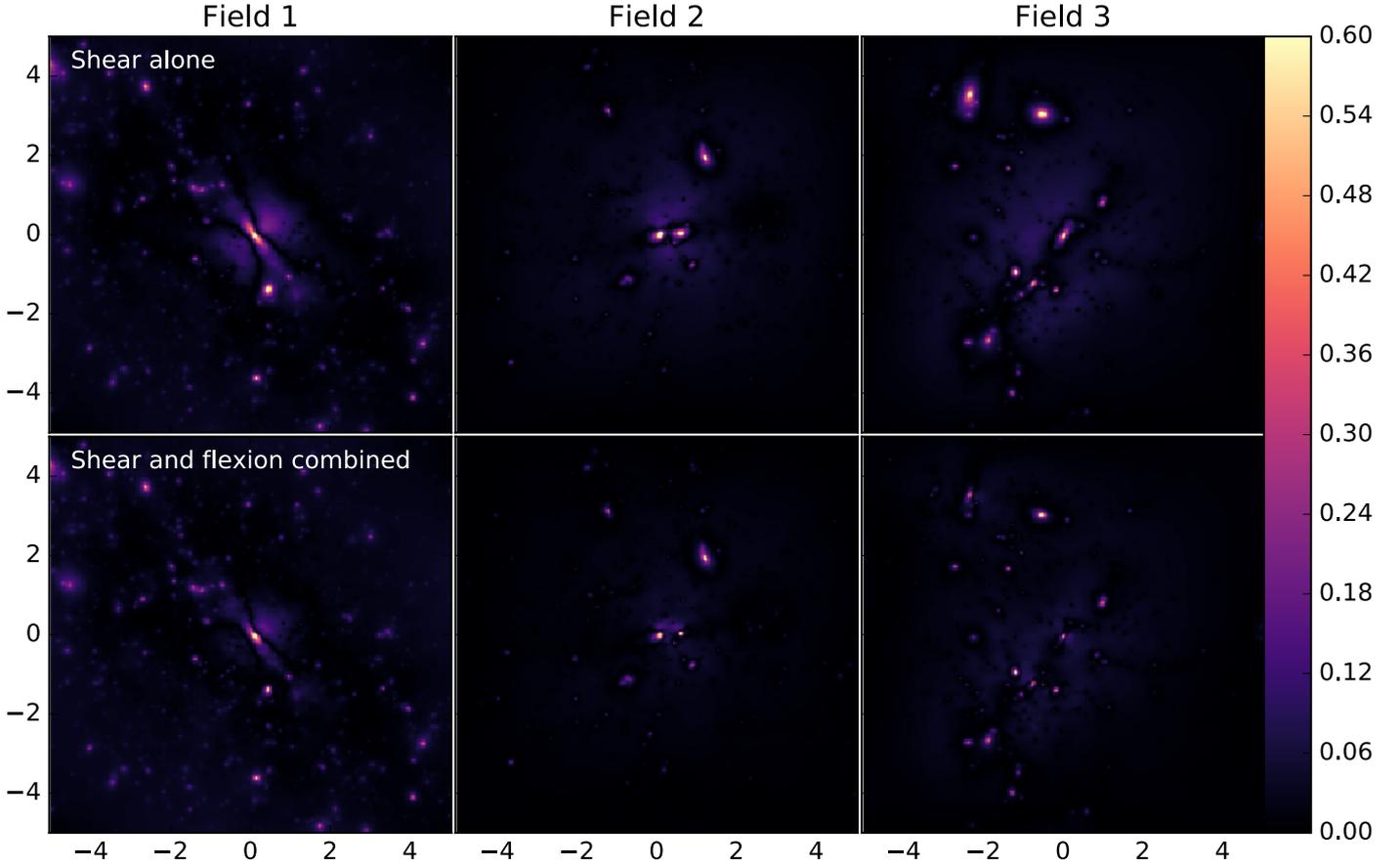


Fig. 11. Absolute difference between the mean of 100 independent signal realisations and the input convergence maps. The *top row* corresponds to reconstructions using the shear alone while the *bottom row* corresponds to reconstructions using shear and flexion.

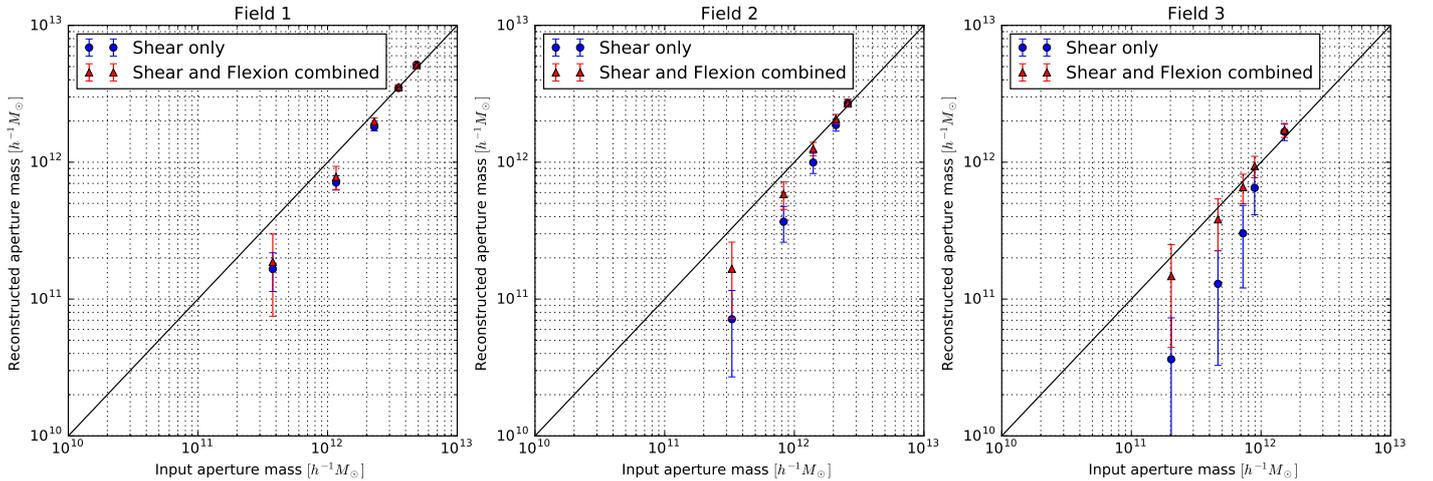


Fig. 12. Aperture masses measured at the centre on the main halo for each field and for five different radii: 19, 38, 75, 150, and 300 h^{-1} kpc or 6'', 12'', 24'', 48'', and 1.6'. The blue circles indicate the masses computed on the shear only reconstructions while the red triangles indicate the masses computed on the combined shear and flexion reconstructions. Error bars represent the standard deviation of 100 independent realisations.

the input aperture masses. As can be seen for all three fields, the aperture masses computed for the largest radius, which is above the arcminute scale coincide between the shear only and combined shear and flexion reconstruction, showing that on large scales the reconstructions are mostly constrained by the shear information. Furthermore, above the arcminute scale, the measured aperture masses are in excellent agreement with the input

masses for all three fields. The impact of flexion becomes obvious on small scales where one can see that the aperture masses are systematically underestimated from the shear only reconstruction while adding flexion can dramatically improve the recovery of small scale details as can be seen for field 3.

To illustrate further the impact of flexion on the recovery of substructure, we also show in [Fig. 13](#) the aperture mass at

Table 3. Integrated mass in the central 1' of each field, corresponding to a radius of $188 h^{-1}$ kpc at the redshift of the lens.

Field	Input $M_{1'}$ $10^{13} h^{-1} M_{\odot}$	Shear only $M_{1'}$ $10^{13} h^{-1} M_{\odot}$	Combined $M_{1'}$ $10^{13} h^{-1} M_{\odot}$
1	18.5	19.1 ± 0.53	18.7 ± 0.54
2	9.48	10.5 ± 0.62	10.2 ± 0.62
3	5.88	6.89 ± 0.83	6.97 ± 0.71

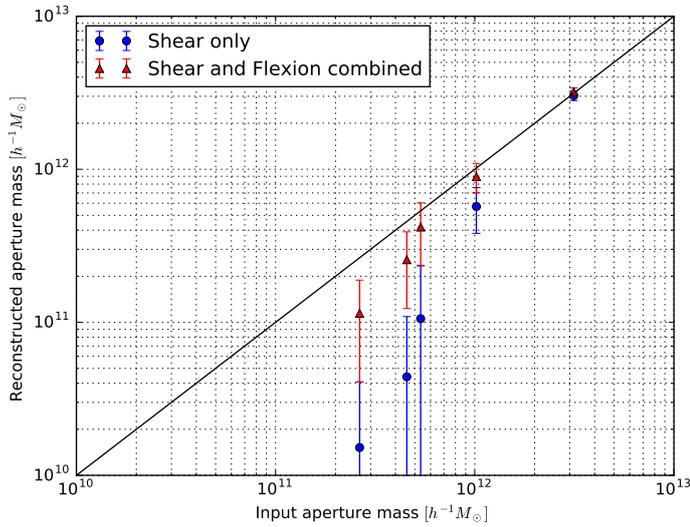


Fig. 13. Aperture mass for the main subhalo in field 1 for five different radii: 19, 38, 75, 150, and $300 h^{-1}$ kpc or 6'', 12'', 24'', 48'', and 1.6'. The blue circles indicate the masses computed on the shear only reconstructions while the red triangles indicate the masses computed on the combined shear and flexion reconstructions. Error bars represent the standard deviation of 100 independent realisations.

the same five radii but for the subhalo visible in the combined shear and flexion reconstruction of field 1. As can be seen in this figure, this substructure, which is around the 15'' scale, is clearly recovered by the addition of flexion between the 12'' and 24'' scale.

6. Conclusion

In this work, we introduced a new mass mapping methodology, aimed at the recovery of high-resolution convergence maps from a combination of shear and flexion measurements. In order to preserve all available small-scale information, our approach avoids any binning or smoothing of the input lensing signal and still relies on efficient Fourier estimators despite the irregular sampling of the shear and flexion fields. Stating the mass mapping problem using Fourier estimators on an irregularly sampled lensing field is an ill-posed linear inverse problem, which we solve using the sparse regularisation framework. Our full reconstruction algorithm incorporates individual redshift, shear and flexion measurements for background galaxies and accounts for the non-linear reduced shear and flexion. We define our sparse regularisation based on a multiscale noise variance map derived from randomisations of the data, which allows us to tune the regularisation by a single parameter corresponding to a significance level for the detection of structures.

We demonstrated the effectiveness of this approach for dealing with irregularly sampled data in the simplified linear case of reconstructing the convergence from noiseless shear measurement. We showed near-perfect recovery of an input convergence field despite 93% of missing data, at a 3'' resolution with only 30 background galaxies per arcmin².

To test our reconstruction algorithm in a realistic setting and, more importantly, to demonstrate the added value of flexion, we built a set of mock catalogues corresponding to a deep HST/ACS survey using as lenses realistic dark matter distributions extracted from N -body simulations. We compared the reconstructions of these simulated fields with and without the inclusion of flexion.

We verified that the shear alone and combined shear and flexion reconstructions agree on scales larger than the arcminute. This behaviour is expected, by construction, since our algorithm weighs the shear and flexion information based on their respective noise levels, promoting shear on large scales and flexion on small scales. More interestingly, we demonstrated that the inclusion of flexion affects the recovery of sub-arcminute scales in two different ways. First and foremost, substructures on the 15'' scale, which are lost when reconstructing from shear alone, can be recovered by the addition of flexion. Second, including flexion allows a much better constraint of the shape of the inner part of the main halos; in particular, we find that anisotropies are more efficiently recovered. These results are in good agreement with the conclusions of Cain et al. (2016), showing the importance of flexion for the recovery of high resolution maps well outside of the inner-most regions of clusters where strong lensing prevails.

In the spirit of reproducible research, our C++ implementation of the mass mapping algorithm presented in this paper will be made freely available at <http://www.cosmostat.org/software/glimpse>. The mock catalogues and their corresponding reconstructions as well as the input convergence maps, are available at http://www.cosmostat.org/product/benchmark_wl.

Acknowledgements. The authors thank Jalal Fadili, Eric Jullo, Fred Ngole, Chieh-An Lin and Martin Kilbinger for useful comments and discussions. This work was funded by the Centre National d'Étude Spatiale (CNES) and the DEDALE project, contract No. 665044, within the H2020 Framework Program of the European Commission.

References

- Bacon, D. J., Goldberg, D. M., Rowe, B. T. P., & Taylor, A. N. 2006, *MNRAS*, 365, 414
- Bartelmann, M. 1995, *A&A*, 303, 643
- Bartelmann, M. 2010, *Class. Quant. Grav.*, 27, 233001
- Beckouche, S., Starck, J. L., & Fadili, J. 2013, *A&A*, 556, A132
- Bradač, M., Lombardi, M., & Schneider, P. 2004, *A&A*, 424, 13
- Bradač, M., Schneider, P., Lombardi, M., & Erben, T. 2005, *A&A*, 437, 39
- Cacciato, M., Bartelmann, M., Meneghetti, M., & Moscardini, L. 2006, *A&A*, 458, 349
- Cain, B., Schechter, P. L., & Bautz, M. W. 2011, *ApJ*, 736, 43
- Cain, B., Bradač, M., & Levinson, R. 2016, *MNRAS*, submitted [[arXiv:1503.08218](https://arxiv.org/abs/1503.08218)]
- Candès, E. J., Wakin, M. B., & Boyd, S. P. 2008, *J. Fourier Analysis and Applications*, 14, 877
- Chambolle, A., & Pock, T. 2011, *J. Mathematical Imaging and Vision*, 40, 120
- Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, *ApJ*, 648, L109
- Condat, L. 2013, *J. Optimization Theory and Applications*, 158, 460
- Deriaz, E., Starck, J. L., & Pires, S. 2012, *A&A*, 540, A34
- Er, X., Li, G., & Schneider, P. 2010, ArXiv e-prints [[arXiv:1008.3088](https://arxiv.org/abs/1008.3088)]
- Goldberg, D. M., & Bacon, D. J. 2005, *ApJ*, 619, 741
- Goldberg, D. M., & Leonard, A. 2007, *ApJ*, 660, 1003

- Jullo, E., Pires, S., Jauzac, M., & Kneib, J. P. 2014, *MNRAS*, **437**, 3969
- Kaiser, N., & Squires, G. 1993, *ApJ*, **404**, 441
- Keiner, J., Kunis, S., & Potts, D. 2009, *ACM Transactions on Mathematical Software*, **36**, 1
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, **740**, 102
- Leonard, A., & King, L. J. 2010, *MNRAS*, **405**, 1854
- Leonard, A., Goldberg, D. M., Haaga, J. L., & Massey, R. 2007, *ApJ*, **666**, 17
- Leonard, A., King, L. J., & Wilkins, S. M. 2009, *MNRAS*, **395**, 1438
- Leonard, A., Dupé, F.-X., & Starck, J.-L. 2012, *A&A*, **539**, A85
- Leonard, A., Lanusse, F., & Starck, J.-L. 2014, *MNRAS*, **440**, 1281
- Massey, R., Williams, L., Smit, R., et al. 2015, *MNRAS*, **449**, 3393
- Merten, J., Cacciato, M., Meneghetti, M., Mignone, C., & Bartelmann, M. 2009, *A&A*, **500**, 681
- Moreau, J.-J. 1962, *CRAS Sér. A Math.*, **255**, 2897
- Okura, Y., Umetsu, K., & Futamase, T. 2007, *ApJ*, **660**, 995
- Paykari, P., Lanusse, F., Starck, J. L., Sureau, F., & Bobin, J. 2014, *A&A*, **566**, A77
- Pires, S., Starck, J. L., Amara, A., et al. 2009, *MNRAS*, **395**, 1265
- Rowe, B., Bacon, D., Massey, R., et al. 2013, *MNRAS*, **435**, 822
- Schneider, P., & Er, X. 2008, *A&A*, **485**, 363
- Seitz, C., & Schneider, P. 1995, *A&A*, **297**, 287
- Seitz, S., & Schneider, P. 1996, *A&A*, **305**, 383
- Starck, J.-L., Murtagh, F., & Fadili, J. 2015, *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis* (Cambridge University Press)
- Starck, J. L., Pires, S., & Refregier, A. 2006, *A&A*, **451**, 1139
- Viola, M., Melchior, P., & Bartelmann, M. 2012, *MNRAS*, **419**, 2215
- Vu, B. C. 2013, *Adv. Comput. Math.*, **38**, 667

Appendix A: Specialisation of the primal-dual algorithm

This appendix introduces elements of proximal calculus and describes the specialisation of the generic primal-dual algorithms introduced in [Condat \(2013\)](#), [Vu \(2013\)](#) to the specific weak lensing mass mapping problems stated in this paper.

A.1. Elements of proximal calculus

Let \mathcal{H} be a finite-dimensional Hilbert space (in the specific case considered in this paper, $\mathcal{H} = \mathbb{R}^n$). Let us consider a function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ and define the *domain of f* , noted $\text{dom } f$, as the subset of \mathcal{H} where f does not reach $+\infty$:

$$\text{dom } f = \{\mathbf{x} \in \mathcal{H} \mid f(\mathbf{x}) < +\infty\}. \quad (\text{A.1})$$

A function f will be said *proper* if its domain $\text{dom } f$ is nonempty. We also define the *epigraph of f* , noted $\text{epi } f$:

$$\text{epi } f = \{(\mathbf{x}, \lambda) \in \mathcal{H} \times \mathbb{R} \mid f(\mathbf{x}) \leq \lambda\}. \quad (\text{A.2})$$

The epigraph is useful to characterise several important properties, in particular the *convexity* and *lower semicontinuity* of f :

$$\text{epi } f \text{ is closed} \Leftrightarrow f \text{ is lower semicontinuous} \quad (\text{A.3})$$

$$\text{epi } f \text{ is convex} \Leftrightarrow f \text{ is convex}. \quad (\text{A.4})$$

We will note Γ_0 the class of proper lower semicontinuous convex functions of \mathbb{R}^n . Functions in Γ_0 are therefore characterised by a non empty closed convex epigraph.

The proximal operator, introduced by [Moreau \(1962\)](#) is at the center of the different algorithms introduced in the following sections. This operator, which can be seen as an extension of the convex projection operator is defined by the following:

Definition 1. Let $F \in \Gamma_0$. For every \mathbf{x} the function $\mathbf{y} \mapsto \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + F(\mathbf{y})$ achieves its infimum at a unique point defined as $\text{prox}_F(\mathbf{x})$.

Therefore, given a proper lower semicontinuous convex function $F \in \Gamma_0$, the proximity operator of F is uniquely defined by:

$$\text{prox}_F(\mathbf{x}) = \arg \min_{\mathbf{y}} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + F(\mathbf{y}). \quad (\text{A.5})$$

In order to illustrate this operator in practice in a simple case, consider $F = i_C$ the indicator function of a closed convex set C . Then the proximity operator reduces to the orthogonal projector onto C , noted proj_C :

$$\text{prox}_{i_C}(\mathbf{x}) = \arg \min_{\mathbf{y} \in C} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 = \text{proj}_C(\mathbf{x}). \quad (\text{A.6})$$

Similarly to the simple case of the indicator function, explicit expressions for the proximity operator exist for a number of different simple functions.

In the context of sparse optimisation, we will be particularly interested in the proximity operator of the ℓ_1 norm. Thankfully, the proximity operator of $F(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$ is explicit and corresponds to *Soft Thresholding*:

$$\text{prox}_{\lambda \|\cdot\|_1}(\mathbf{x}) = \text{ST}_{\lambda}(\mathbf{x}), \quad (\text{A.7})$$

where the Soft Thresholding operator ST_{λ} is defined for $\mathbf{x} \in \mathbb{R}^N$ as:

$$\forall i \in \llbracket 1, N \rrbracket, \quad \text{ST}_{\lambda}(\mathbf{x})_i = \begin{cases} x_i - \lambda & \text{if } x_i \geq \lambda \\ 0 & \text{if } |x_i| \leq \lambda \\ x_i + \lambda & \text{if } x_i \leq -\lambda. \end{cases} \quad (\text{A.8})$$

As previously mentioned, the proximity operator is a key tool in convex optimisation as it can be used to derive fast convergent algorithms for convex but not necessarily differentiable problems. In particular, in this paper we consider optimisation problems of the following general form:

$$\arg \min_{\mathbf{x}} F(\mathbf{x}) + G(W(\mathbf{x})) + H(\mathbf{x}), \quad (\text{A.9})$$

where F is convex and differentiable, with a Lipschitzian gradient of constant β , $(H, G) \in \Gamma_0^2$ and W is a non-zero linear operator. In our case, F will typically be the χ^2 data fidelity term, W the wavelet transform, and G and H the ℓ_1 norm and the indicator function respectively, which are convex but not differentiable. Proximal algorithm generally rely on an iterative scheme using the gradient of the first term and the proximity operators of the additional terms. The particular difficulty of this problem stems from the composition of the two operators G and W which does not possess in the general case an explicit proximity operator even if prox_G has a closed-form expression.

Primal-dual algorithms, such as the one initially proposed in [Chambolle & Pock \(2011\)](#), avoid this difficulty by recasting part of the problem in a form which allows the separation between G and W such that a closed-form expression of prox_G can be used if it exists. The main limitation of this algorithm however is that it does not take advantage of the differentiability of the first term F . More recently, a variation of this approach was proposed in [Condat \(2013\)](#), [Vu \(2013\)](#), still relying on a primal-dual formulation for separating G and W but taking full advantage of the differentiability of F , thus leading to a much more efficient algorithm in practice. The main iteration of this minimisation algorithm for solving Eq. (A.9) takes the following form:

$$\begin{cases} \mathbf{x}^{(k+1)} &= \text{prox}_{\tau H}(\mathbf{x}^{(k)} + \tau(\nabla F(\mathbf{x}^{(k)}) - W^* \mathbf{u}^{(k)})) \\ \mathbf{u}^{(k+1)} &= \text{prox}_{G^*}(\mathbf{u}^{(k)} + \sigma W(2\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})) \end{cases}, \quad (\text{A.10})$$

where τ and σ must verify $1 - \tau\sigma \|W\|^2 > \tau\beta/2$ to ensure convergence and where G^* is the convex conjugate of the original function G , defined according to $G^*(\mathbf{x}) = \max_{\mathbf{y}} \{\langle \mathbf{y}, \mathbf{x} \rangle - G(\mathbf{y})\}$. Thankfully, the proximal operator of the convex conjugate is readily expressible in terms of the proximal operator of the original function, with $\text{prox}_{G^*}(\mathbf{x}) = \mathbf{x} - \text{prox}_G(\mathbf{x})$, so that the proximity operator of the conjugate can be evaluated at no extra-cost.

A.2. Algorithm specialisation in the simplified linear setting

In the simplified linear setting, deriving the specialisation of minimisation algorithm introduced in Eq. (A.10) is straightforward. We aim to solve Eq. (37), which can be seen as a specialisation of Eq. (A.9) for the following choice of operators:

$$\begin{aligned} F(\boldsymbol{\kappa}) &= \frac{1}{2} \|\Sigma_{\gamma}^{-\frac{1}{2}} [\boldsymbol{\gamma} - \mathbf{TPF}^* \boldsymbol{\kappa}]\|_2^2; & G(\boldsymbol{\alpha}) &= \lambda \|\mathbf{w} \circ \boldsymbol{\alpha}\|_1 \\ W(\boldsymbol{\kappa}) &= \boldsymbol{\Phi}^* \boldsymbol{\kappa}; & H(\boldsymbol{\kappa}) &= i_{\mathcal{S}(\cdot)=0}(\boldsymbol{\kappa}). \end{aligned}$$

Applying the [Condat \(2013\)](#), [Vu \(2013\)](#) algorithm introduced in the previous section is therefore only a matter of computing the

gradient of F and the proximity operators of G^* and H . The gradient of the quadratic term F is simply:

$$\nabla F(\boldsymbol{\kappa}) = \mathbf{FP}^* \mathbf{T}^* \Sigma_{\gamma}^{-1} (\boldsymbol{\gamma} - \mathbf{TPF}^* \boldsymbol{\kappa}), \quad (\text{A.11})$$

while the proximity operators of G and H have a closed-form expression already introduced in the previous section:

$$\text{prox}_{G^*}(\boldsymbol{\alpha}) = \boldsymbol{\alpha} - \text{ST}_{\lambda_{\text{ow}}}(\boldsymbol{\alpha}), \quad (\text{A.12})$$

$$\text{prox}_H(\boldsymbol{\kappa}) = \mathfrak{X}(\boldsymbol{\kappa}). \quad (\text{A.13})$$

These expressions for the gradient and proximity operators readily yield [algorithm 1](#) when applied to the iteration introduced in Eq. (A.10).

A.3. Algorithm specialisation for the complete problem

Contrary to the previous case, solving the full mass mapping problem stated in Eq. (52) presents a number of difficulties. The specialisation of Eq. (A.9) differs from the previous case in the definition of the differentiable term F and the convex function H :

$$F\left(\begin{bmatrix} \boldsymbol{\kappa} \\ \tilde{\mathcal{F}} \end{bmatrix}\right) = \frac{1}{2} \|C_{\kappa g}^{-1} [(1 - \mathbf{ZTF}^* \boldsymbol{\kappa})\boldsymbol{g} - \mathbf{ZTPF}^* \boldsymbol{\kappa}]\|_2^2 \\ + \frac{1}{2} \|C_{\kappa F}^{-1} [(1 - \mathbf{ZTF}^* \boldsymbol{\kappa})\boldsymbol{F} - \mathbf{ZTF}^* \tilde{\mathcal{F}}]\|_2^2,$$

$$H\left(\begin{bmatrix} \boldsymbol{\kappa} \\ \tilde{\mathcal{F}} \end{bmatrix}\right) = \text{i}_{\text{Im}(\mathbf{R})}\left(\begin{bmatrix} \boldsymbol{\kappa} \\ \tilde{\mathcal{F}} \end{bmatrix}\right).$$

Remember that this expression for the quadratic fidelity term F is a linearisation of the original problem which includes a non-linear factor due to the reduced shear. However, despite this first simplification, the gradient of this expression proves to be problematic. Since the shear and flexion terms are equivalent and separable, let us concentrate on the gradient of the shear term of F , which takes the form:

$$\mathbf{FP}^* \mathbf{T}^* \mathbf{Z} C_{\kappa g}^{-2} [(1 - \mathbf{ZTF}^* \boldsymbol{\kappa})\boldsymbol{g} - \mathbf{ZTPF}^* \boldsymbol{\kappa}] \\ + \mathbf{FT}^* \mathbf{Z} \boldsymbol{g}^* C_{\kappa g}^{-2} [(1 - \mathbf{ZTF}^* \boldsymbol{\kappa})\boldsymbol{g} - \mathbf{ZTPF}^* \boldsymbol{\kappa}]. \quad (\text{A.14})$$

The first term of this expression simply corresponds to the same gradient as in the linear case but corrected for the reduced shear. In particular, the noise affecting the measured shear still propagates linearly. On the contrary, the second term is new compared to the linear case and is problematic in the sense that it becomes a quadratic function of the reduced shear \boldsymbol{g} . If one assumes the reduced shear noise to be Gaussian, the noise contribution of this second term becomes χ^2 distributed. As explained in [Sect. 3.4](#), our regularisation scheme is defined in terms of the

standard deviation of the noise propagated to the wavelet coefficients. While this scheme proves very effective for Gaussian noise, it is not appropriate for χ^2 distributed noise.

Therefore, in the approach presented here, we choose to use a suboptimal gradient for the quadratic data fidelity term by keeping only the first term of Eq. (A.14), which is the price to pay in order to keep the very effective regularisation scheme introduced in the linear case. Thus, we use the following expression instead of the exact gradient of F :

$$\nabla F = \mathbf{FP}^* \mathbf{T}^* \mathbf{Z} C_{\kappa g}^{-2} ((1 - \mathbf{ZTF}^* \boldsymbol{\kappa})\boldsymbol{g} - \mathbf{ZTPF}^* \boldsymbol{\kappa}) \\ + \mathbf{FT}^* \mathbf{Z} C_{\kappa F}^{-2} ((1 - \mathbf{ZTF}^* \boldsymbol{\kappa})\boldsymbol{F} - \mathbf{ZTF}^* \tilde{\mathcal{F}}).$$

Note however that despite this simplification, the resulting procedure still corresponds to the reduced shear correction scheme suggested in [Seitz & Schneider \(1995\)](#).

The second difference with the shear only reconstruction comes from the more complex constraint H based on the indicator function of the image of the linear operator \mathbf{R} introduced in Eq. (51). Thankfully, despite the additional complexity, the proximity operator of this term can still be explicitly from its definition:

$$\text{prox}_{\text{i}_{\text{Im}(\mathbf{R})}}\left(\begin{bmatrix} \boldsymbol{\kappa} \\ \tilde{\mathcal{F}} \end{bmatrix}\right) = \arg \min_{(\boldsymbol{x}_{\kappa}, \boldsymbol{x}_{\mathcal{F}})} \frac{1}{2} \left\| \begin{bmatrix} \boldsymbol{\kappa} \\ \tilde{\mathcal{F}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{x}_{\kappa} \\ \boldsymbol{x}_{\mathcal{F}} \end{bmatrix} \right\|_2^2 + \text{i}_{\text{Im}(\mathbf{R})}\left(\begin{bmatrix} \boldsymbol{x}_{\kappa} \\ \boldsymbol{x}_{\mathcal{F}} \end{bmatrix}\right) \\ = \mathbf{R} \arg \min_{\boldsymbol{x}_{\kappa} \in \mathbb{R}^{N \times N}} \left\| \begin{bmatrix} \boldsymbol{\kappa} \\ \tilde{\mathcal{F}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{x}_{\kappa} \\ \mathbf{F} \mathbf{Q} \mathbf{F}^* \boldsymbol{x}_{\kappa} \end{bmatrix} \right\|_2^2 \\ = \mathbf{R} \mathfrak{X} \left(\mathbf{F} \frac{1}{k^2 + \frac{\sigma_{\mathcal{F}}^2}{\sigma_{\epsilon}^2}} \left(k^2 \mathbf{Q}^{-1} \mathbf{F}^* \tilde{\mathcal{F}} + \frac{\sigma_{\mathcal{F}}^2}{\sigma_{\epsilon}^2} \mathbf{F}^* \boldsymbol{\kappa} \right) \right).$$

In this expression, σ_{ϵ}^2 is the variance of the intrinsic ellipticity and $\sigma_{\mathcal{F}}^2$ is the variance of the intrinsic flexion. Let us detail what is computed by this proximity operator. The first step is to rewrite the proximity operator using its definition, introduced in Eq. (A.5). Then, using the definition of $\text{Im}(\mathbf{R})$, the problem can be recast in terms of a single real unknown \boldsymbol{x}_{κ} , from which convergence and flexion can be computed by applying the operator \mathbf{R} . Finally, as $\boldsymbol{\kappa}$ is fitted to the shear and $\tilde{\mathcal{F}}$ is fitted to the flexion, solving the minimisation problem for \boldsymbol{x}_{κ} amounts to finding the minimum variance filter combining shear and flexion, already introduced in Eq. (17). To summarise, this operator computes the optimal combination of the two input variables, making explicit use of the flexion Fourier operator and its inverse $\mathbf{Q}^{-1} = \mathbf{Q}^*/k^2$, thus eliminating the need to solve this additional problem as part of the main minimisation problem.

Given these expressions for ∇F and $\text{prox}_{\text{i}_{\text{Im}(\mathbf{R})}}$, the specialisation of the primal-dual algorithm yields [algorithm 2](#).