

## Open Access Book Chapter

### Published version

Chapter title	Exploring the Components of the Universe Through Higher-Order Weak Lensing Statistics
Title in English	
Author(s)	Adrienne Leonard, Jean-Luc Starck, Sandrine Pires and François-Xavier Dupé
DOI	10.5772/51871
Copyright holder	<b>The Authors</b>
License	CC BY 3.0 license
Language	English

### Original publication

Book title	Open Questions in Cosmology
Title in English	
Publisher	<b>Intech</b>
Publication date	2012
ISBN	<b>9789535108801</b>
DOI	
Editor(s)	Gonzalo J. Olmo

### ERC Grant information

Project title	EC   SPARSEASTRO
Project acronym	SparseAstro
Grant number	228261
ERC principal investigator	Jean-Luc Starck



---

# Exploring the Components of the Universe Through Higher-Order Weak Lensing Statistics

---

Adrienne Leonard, Jean-Luc Starck,  
Sandrine Pires and François-Xavier Dupé

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/51871>

---

## 1. Introduction

Our current cosmological model, backed by a large body of evidence from a variety of different cosmological probes (for example, see [1, 2]), describes a Universe comprised of around 5% normal baryonic matter, 22% cold dark matter and 73% dark energy. While many cosmologists accept this so-called concordance cosmology – the  $\Lambda$ CDM cosmological model – as accurate, very little is known about the nature and properties of these dark components of the Universe.

Studies of the cosmic microwave background (CMB), combined with other observational evidence of big bang nucleosynthesis indicate that dark matter is non-baryonic. This supports measurements on galaxy and cluster scales, which found evidence of a large proportion of dark matter. This dark matter appears to be cold and collisionless, apparent only through its gravitational effects.

While dark matter is largely responsible for the growth of the largest structures in the Universe, dark energy – dominant at late times – appears to have a negative pressure, and to be responsible for an accelerated expansion of the Universe [3]. It is usually parameterised by its equation of state parameter  $w = p/\rho$ , where  $p$  is the pressure associated with the dark energy and  $\rho$  is its energy density. An equation of state parameter of  $w = -1$  would indicate a cosmological constant, consistent with general relativistic theory, but deviations from this value would suggest a rather more exotic dark energy, and might perhaps imply a modification to our current theory of gravity.

Gravitational lensing – the deflection of light rays by massive objects due to gravitational effects, which gives rise to distortions in images of galaxies – is an ideal probe of the dark universe, as it can probe the evolution of the dark matter power spectrum in an unbiased way, and offers complementary constraints to those obtained from the CMB and other probes of large-scale structure.

Moreover, gravitational lensing probes cosmological perturbations on smaller angular scales than CMB studies, and is thus sensitive to the non-Gaussianity induced by the late-time non-linear evolution of structures such as clusters of galaxies, as well as any primordial non-Gaussianity arising, for example, due to inflation very early in the evolution of the Universe.

Traditionally, constraints from gravitational lensing are obtained by considering two-point statistics of the lensing shear field, which encodes the small elliptical distortions applied to the images of galaxies as a result of the gravitational potential field of structures along the light's path. Such two-point statistics are only weakly sensitive to the dark energy density parameter  $\Omega_\Lambda$ <sup>1</sup>, and depend on a degenerate combination of the amplitude of the matter power spectrum  $\sigma_8$  and the matter density parameter  $\Omega_M$ . In addition, two-point statistics probe only the Gaussian part of the shear field, therefore in considering such statistics alone, information about nonlinear structure evolution and primordial non-Gaussianity is lost.

Similarly, attempts to reconstruct a map of the two-dimensional projected surface mass density (the convergence  $\kappa$ ) and three dimensional density field have often involved the use of Gaussian priors to constrain the reconstruction (for example, see [4]), thus again having limited application in studies of non-Gaussianity.

In this chapter, we present a review of recently developed gravitational lensing techniques that go beyond the standard two-point statistics, both in the arena of map-making in two and three dimensions and that of higher-order statistics of the shear field or, equivalently, the convergence field  $\kappa$ .

The methods we present are all based on the concept of sparse recovery, which has been found to be a very powerful tool in signal processing [5, 6]. Such methods are based on the assumption that a given image or observation can be represented sparsely in an appropriate basis (such as a Fourier or wavelet basis). Sparse priors using a wavelet basis have been used in many areas of signal processing in astronomy; of particular interest in this chapter will be the areas of denoising and in the reconstruction of the 3D density field from lensing measurements, but other applications include deconvolution and inpainting.

There is much information to be gained by considering non-Gaussian statistics and nonlinear signal processing methods, and this is an exciting and active area of research. With new surveys such as the Euclid satellite [7] coming online within the next decade, a wealth of high-quality data will soon be available. These new techniques will therefore prove invaluable in constraining the cosmological model, and allow us to better understand the nature of the primary constituents of the Universe.

## 2. Weak gravitational lensing

We begin with a brief overview of weak lensing theory. For a more complete description and discussion of the subject, see [8, 9]

### 2.1. Weak lensing theory

The basic idea underlying the theory of gravitational lensing is that massive objects distort the spacetime around them, and thus bend the path of light in their vicinity. A light ray

<sup>1</sup> The energy density parameters are defined by  $\Omega_i = \rho_i / \rho_c$ , where  $\rho_c$  is the critical density of the Universe

originating at angle  $\beta$  is deflected such that it appears to the observer to originate at an angular position  $\theta$ , where

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

where  $\psi$  is the two-dimensional deflection potential associated with the lens and  $\alpha$  is the deflection angle.

For extended sources, photons from different angular positions in the source plane are deflected differently, giving rise to a distortion in the observed galaxy image, which is described – to first order – by the Jacobian of the lens equation (1):

$$\mathcal{A} = \frac{\partial\beta}{\partial\theta} = \delta_{i,j} - \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (2)$$

where  $\kappa = \frac{1}{2}\nabla^2\psi$  is the convergence, or dimensionless surface density, and  $\gamma = \gamma_1 + i\gamma_2 = |\gamma|e^{2i\phi}$  is the complex shear, which gives rise to an anisotropic elliptical distortion in the lensed image. The components of the shear are related to the potential via:

$$\gamma_1 = \frac{1}{2}(\partial_1^2\psi - \partial_2^2\psi), \quad \gamma_2 = \partial_1\partial_2\psi, \quad (3)$$

where  $\partial_i = \partial/\partial\theta_i$ , and the shear is related to the convergence through the relation:

$$\gamma(\theta) = \frac{1}{\pi} \int d^2\theta' \mathcal{D}(\theta - \theta') \kappa(\theta'), \quad (4)$$

where the convolution kernel  $\mathcal{D}$  is given by

$$\mathcal{D}(\theta) \equiv \frac{\theta_2^2 - \theta_1^2 - 2i\theta_1\theta_2}{|\theta|^4} = -\frac{1}{(\theta_1 - i\theta_2)^2}. \quad (5)$$

The convergence, in turn, can be related to the 3D density contrast  $\delta(\mathbf{r}) \equiv \rho(\mathbf{r})/\bar{\rho} - 1$  by

$$\kappa(\theta, w) = \frac{3H_0^2\Omega_M}{2c^2} \int_0^w dw' \frac{f_K(w')f_K(w-w')}{f_K(w)} \frac{\delta[f_K(w')\theta, w']}{a(w')}, \quad (6)$$

where  $\bar{\rho}$  is the mean density of the Universe,  $H_0$  is the Hubble parameter,  $\Omega_M$  is the matter density parameter,  $c$  is the speed of light,  $a(w)$  is the scale parameter evaluated at comoving distance  $w$ , and

$$f_K(w) = \begin{cases} K^{-1/2} \sin(K^{1/2}w), & K > 0 \\ w, & K = 0 \\ (-K)^{-1/2} \sinh([-K]^{1/2}w) & K < 0 \end{cases}, \quad (7)$$

gives the comoving angular diameter distance as a function of the comoving distance and the curvature,  $K$ , of the Universe.

## 2.2. Two-point statistics

The typical gravitational shear applied to a galaxy as a result of lensing by large-scale structure in the Universe – so-called cosmic shear – is of order  $|\gamma| \sim 0.01$ . However, galaxies are intrinsically elliptical in shape, with a typical ellipticity of order  $|\varepsilon| \sim 0.2 - 0.3$ , therefore the gravitational lensing effect can only be measured statistically; under the assumption that galaxy shapes are intrinsically uncorrelated, the mean ellipticity computed over a large number of sources will yield the gravitational shear:  $\langle \varepsilon \rangle \simeq \gamma$ .

The most common method for constraining cosmological parameters in weak lensing studies is to use two-point statistics of the shear field. The power spectrum of the shear or convergence,  $P_\kappa(\ell) = P_\gamma(\ell)$ , can be related directly to the 3D matter power spectrum of density fluctuations  $\delta$  by:

$$P_\kappa(\ell) = \frac{9H_0^4 \Omega_M^2}{4c^4} \int dw \frac{W^2(w)}{a^2(w)} P_\delta \left( \frac{\ell}{f_K(w)}, w \right), \quad (8)$$

where

$$W(w) = \int_w^\infty dw' \frac{f_K(w' - w)}{f_K(w')} p(z) \left[ \frac{dz}{dw'} \right], \quad (9)$$

is a weighting function integrated over the probability distribution of sources  $p(z)$  in the sample as a function of redshift  $z$ .

When working with real data, it is often more convenient to consider statistics computed in real space, namely the shear correlation functions, which are defined in terms of the convergence power spectrum as [10, 11]:

$$\zeta_\pm(\theta) = \langle \gamma_t \gamma_t \rangle \pm \langle \gamma_\times \gamma_\times \rangle = \int_0^\infty \frac{d\ell}{2\pi} J_{0,A}(\ell\theta) P_\kappa(\ell), \quad (10)$$

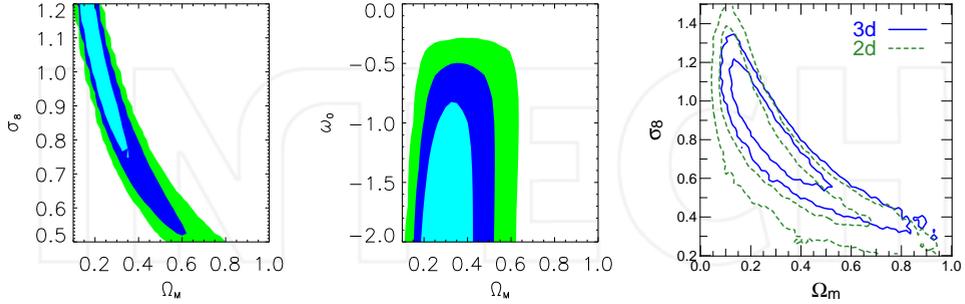
where  $J_n$  is an  $n$ -th order Bessel function of the first kind, and  $\gamma_t$  and  $\gamma_\times$  are the tangential and cross shear components, respectively, which are defined relative to the vector connecting the two galaxies separated by angular distance  $\theta$ .

## 2.3. Cosmological constraints

The measured two-point statistics described above can, by virtue of their relationship to the power spectrum of the underlying matter density fluctuations, be used to place constraints on cosmological parameters. These constraints may be improved if information about the distances to source galaxies is used to bin the galaxies into redshift slices, and compute the correlation functions tomographically.

Figure 1 shows cosmological parameter constraints obtained using two state-of-the-art telescopes: the Canada-France-Hawaii Telescope (CFHT), and the Hubble Space Telescope (specifically, the Cosmos survey). While the CFHT results were obtained using a 2D analysis [12], the Cosmos results show constraints obtained by both a 2D and a 3D analysis of

the shear data [13]. In both cases, a large degeneracy is seen between the matter density parameter  $\Omega_M$  and  $\sigma_8$ , the amplitude of the matter power spectrum. The CFHT results also show constraints on the dark energy equation of state parameter  $w$ . Here, again, a degeneracy between parameters is seen.



(a) Joint constraints on  $\Omega_M$  and  $\sigma_8$  from a 2D analysis of the combined CFHT wide and deep surveys [12]. (b) Joint constraints on  $\Omega_M$  and  $w$  from a 2D analysis of the CFHT deep survey [12]. (c) Joint constraints on  $\Omega_M$  and  $\sigma_8$  from a 2D and 3D analysis of the Cosmos survey data [13].

**Figure 1.** Cosmological parameter constraints from recent weak lensing surveys. Contours are plotted at the 68% (cyan), 95% (blue) and 99.9% (green) confidence levels for the CFHT data, and at the 68% (inner) and 95% (outer) confidence levels for the Cosmos data.

While it is clear from the figures that two-point shear statistics hold a wealth of information, it is also evident that these statistics depend on degenerate combinations of the cosmological model parameters. In order to break these degeneracies, and thus more tightly constrain our cosmological model, we must therefore consider higher-order statistics of the shear or convergence. Such higher-order statistics will enable us to capture information about the non-Gaussian part of the lensing spectrum due to a combination of primordial non-Gaussianity and late-time nonlinear evolution of structures, which two-point statistics are unable to probe.

### 3. Scalar fields associated with the shear

The shear  $\gamma$  is a spin-2 field, therefore while the two-point correlations of the shear field can be reduced to a scalar quantity for parity reasons, this is not the case for higher-order moments of the shear field [14–16]. For this reason, it is useful to compute a scalar quantity from the shear field before computing higher-order statistics.

#### 3.1. The aperture mass statistic

For this purpose, the aperture mass statistic  $M_{ap}$  [17, 18] is widely-used. The  $M_{ap}$  statistic is defined as the convolution of the convergence  $\kappa$  with a radially symmetrical filter function  $U(|\vartheta|)$  of width  $\theta$ :

$$M_{ap}(\theta) = \int d^2\vartheta \kappa(\vartheta)U(|\vartheta|) . \tag{11}$$

By considering the relationship between the shear,  $\gamma$ , and the convergence, one can reformulate equation (11) in terms of the measured shear as

$$M_{ap}(\boldsymbol{\theta}) = \int d^2\boldsymbol{\vartheta} \gamma_t(\boldsymbol{\vartheta}) Q(|\boldsymbol{\vartheta}|), \quad (12)$$

where  $\gamma_t(\boldsymbol{\vartheta})$  is the tangential component of the shear at position  $\boldsymbol{\vartheta}$  relative to the centre of the aperture,  $Q(|\boldsymbol{\vartheta}|)$  is a second radially-symmetric function, related to  $U(|\boldsymbol{\vartheta}|)$  by:

$$Q(\vartheta) \equiv \frac{2}{\vartheta^2} \int_0^\vartheta \vartheta' U(\vartheta') d\vartheta' - U(\vartheta), \quad (13)$$

and  $U(\vartheta)$  is required to be compensated, i.e.

$$\int_0^{\vartheta_{cut}} \vartheta U(\vartheta) d\vartheta = 0, \quad (14)$$

with  $\vartheta_{cut}$  often taken to be the radius of the aperture,  $\theta$ . Furthermore,  $Q(\vartheta)$  and  $U(\vartheta)$  are required to go to zero smoothly at  $\vartheta_{cut}$ . It is also preferable that the power spectrum of  $U(\vartheta)$  is local in the frequency domain, and shows no oscillatory behaviour. This ensures that the filter function acts as a band-pass filter, allowing detection of structures at the scale of interest only.

Several authors [19, 20] have advocated a filter function of the form:

$$\begin{aligned} U(\vartheta) &= \frac{A}{\vartheta^2} \left( 1 - \frac{b\vartheta^2}{\vartheta^2} \right) \exp\left(-\frac{b\vartheta^2}{\vartheta^2}\right), \\ Q(\vartheta) &= \frac{A}{\vartheta^2} \frac{b\vartheta^2}{\vartheta^2} \exp\left(-\frac{b\vartheta^2}{\vartheta^2}\right), \end{aligned} \quad (15)$$

where various choices for the constant  $b$  and the overall normalisation  $A$  have been used in the literature [20–22]. Of specific interest is the form used in [20] and [22], where  $b = 1/2$  is chosen.<sup>2</sup> This form is considered to be optimal for higher-order weak lensing statistics, such as the skewness of the aperture mass statistic [19, 22].

Note that the  $Q(\vartheta)$  filter function described in equation (15) shows a peak at  $\vartheta = \sqrt{2}\theta$ , and tends to zero as  $\vartheta \rightarrow \infty$ . In practice, any algorithm used to generate an aperture mass map will need to truncate this filter function at some finite radius. This will involve a trade-off between accuracy and algorithm speed, and truncation may affect the effective width of the filter.

One advantage of this method is that the aperture mass statistic can be computed directly from the shear catalogue. Moreover, as the filter acts as a band-pass filter, it is possible to boost the signal relative to the noise by considering a filter with a scale substantially larger

<sup>2</sup> This form is analogous to the Mexican Hat wavelet function

than the typical scale of the noise (which is typically dominant on pixel scales, depending on the binning chosen for the shear data). Indeed, optimal signal-to-noise is obtained when the filter is chosen such that its angular scale matches that of the structures we aim to detect, and its shape matches closely the expected profile of these structures.

### 3.2. The wavelet transform

The wavelet transform is a multiscale transform, where the wavelet coefficients of an image are computed at each position in the image at various different scales simultaneously. In one dimension, the wavelet coefficient of a function  $f(x)$ , evaluated at position  $b$  and scale  $a$  is defined as [23, 24]:

$$W(a, b) = \frac{1}{\sqrt{a}} \int f(x) \psi^* \left( \frac{x-b}{a} \right) dx, \quad (16)$$

where  $\psi(x)$  is the analysing wavelet. The analysis is analogous in 2 dimensions, with  $\psi(x, y) = \psi(x)\psi(y)$ .

By definition, wavelets are compensated functions; i.e. the wavelet function  $\psi(x)$  is constrained such that  $\int_{\mathbb{R}^1} \psi(x) dx = 0$  and hence, by extension

$$\iint_{\mathbb{R}^2} \psi(x, y) dx dy = 0. \quad (17)$$

According to the definition in equation (16), the continuous wavelet transform of an image is therefore nothing more than the convolution of that image with compensated filter functions of various characteristic scales. If the image  $f(x, y)$  is taken to be the convergence  $\kappa(x, y)$ , then for an appropriate choice of (radially-symmetric, local) wavelet, the wavelet transform is formally identical to the aperture mass statistic at the corresponding scales, the only difference being the choice of filter functions.

In practice in application, we use the starlet transform algorithm [6, 23–25], which simultaneously computes the wavelet transform on dyadic scales corresponding to  $2^j$  pixels. This algorithm decomposes the convergence map of size  $N \times N$  into  $J = j_{\max} + 1$  sub-arrays of size  $N \times N$  as follows:

$$\kappa(x, y) = c_J(x, y) + \sum_{j=1}^{j_{\max}} w_j(x, y), \quad (18)$$

where  $j_{\max}$  represents the number of wavelet bands (or, equivalently, aperture mass maps) considered,  $c_J$  represents a smooth (or continuum) version of the original image  $\kappa$ , and  $w_j$  represents the input map filtered at scale  $2^j$  (i.e. the aperture mass map at  $\theta = 2^j$  pixels).

Using the wavelet formalism to derive the aperture mass statistic presents a number of advantages. Many families of wavelet functions have been studied in the statistical literature, and all these wavelet functions could be applied to the aperture mass statistic. This allows us to tune our filter function to optimise the signal-to-noise in the resulting maps. In addition, for some specific wavelet functions, discrete and very fast algorithms exist, allowing us to

compute a set of wavelet scales through the use of a filter bank with a very limited number of operations. See [6] for a full review of the different wavelet transform algorithms.

In the starlet transform algorithm, the wavelet  $\psi(x, y)$  is separable and can be defined by:

$$\psi\left(\frac{x}{2}, \frac{y}{2}\right) = 4 \left[ \varphi(x, y) - \frac{1}{4} \varphi\left(\frac{x}{2}, \frac{y}{2}\right) \right], \quad (19)$$

where  $\varphi(x, y) = \varphi(x)\varphi(y)$  and  $\varphi(x)$  is a scaling function from which the wavelet is generated. In the case of the starlet wavelet,  $\varphi(x)$  is a B3-spline:

$$\varphi(x) = \frac{1}{12} (|x-2|^3 - 4|x-1|^3 + 6|x|^3 - 4|x+1|^3 + |x+2|^3), \quad (20)$$

which is a compact function that is identically zero for  $|x| > 2$ .

This wavelet function has a compact support in real space, is well localized in Fourier domain, and the wavelet decomposition of an image can be obtained with a very fast algorithm (see [24] for a full description).

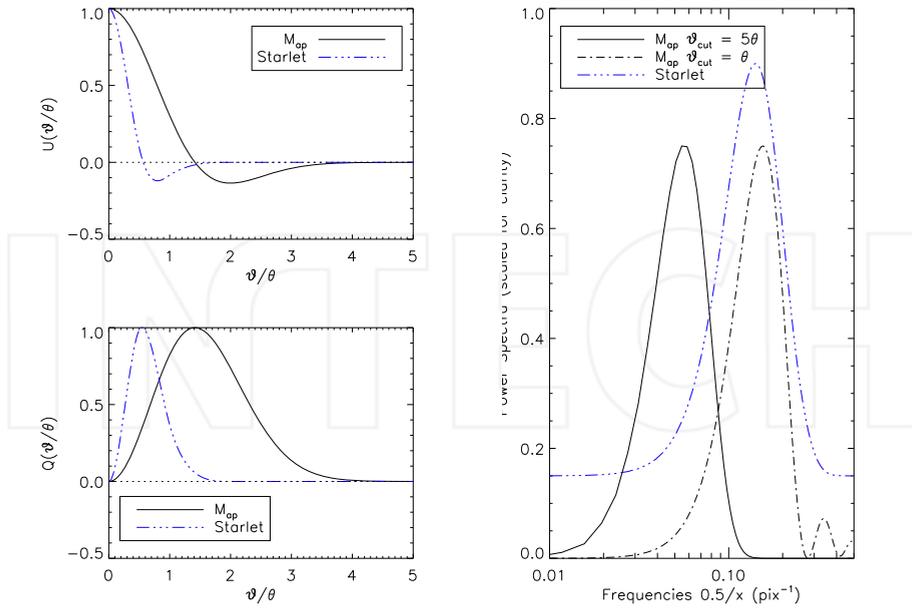
Figure 2 shows the starlet wavelet and aperture mass filters defined above in both real and Fourier space. Notice that the two filters presented have very similar shapes in real space, but different widths. While the starlet filter function goes to zero identically at  $\vartheta = 2\theta$ , and remains zero beyond this value, the  $M_{ap}$  filter function goes to zero as  $\vartheta \rightarrow \infty$ , and must therefore be truncated when applied in practice. Clearly at  $\vartheta = 5\theta$ , the  $M_{ap}$  filters are sufficiently close to zero, so that this is an appropriate truncation radius; however this will impact the computation time of the  $M_{ap}$  statistic.

In Fourier space, we show the response of the  $M_{ap}$  filter for truncation radii of  $\vartheta_{cut} = \theta$  and  $5\theta$ . Notice that the shape of the response curves in Fourier space for both the  $M_{ap}$  and wavelet filter functions are similar, but that the peak of the response curve for the  $M_{ap}$  filter function shifts when  $\vartheta_{cut}$  is varied. This is because the effective scale of the filter function is being changed. Also notice that in the case of  $\vartheta_{cut} = \theta$ , high-frequency oscillations are present in the response curve. This is a direct result of the truncation of the filter function, and will occur whenever such truncation is applied. Indeed, imperceptibly small oscillations are still present in the response curve for the filter truncated at  $\vartheta_{cut} = 5\theta$ .

Such oscillations are not present at all with the starlet filter function, as no truncation is applied to this function whatsoever. This gives the starlet transform the distinct advantage of being directly applicable, without any consideration needed regarding truncation radii and the associated impact on the Fourier-space response of the filter functions.

### 3.3. Advantages of the wavelet formalism

The aperture mass formalism tends to be the preferred method for weak lensing studies for a number of reasons. Firstly, the filter functions can be simply expressed analytically, and any associated statistics of the aperture mass can therefore be straightforwardly computed from



(a)  $M_{ap}$  and wavelet filter functions in real space

(b) Filter response curves in Fourier space

**Figure 2.** A comparison of the  $M_{ap}$  and starlet wavelet filters in real and Fourier space.

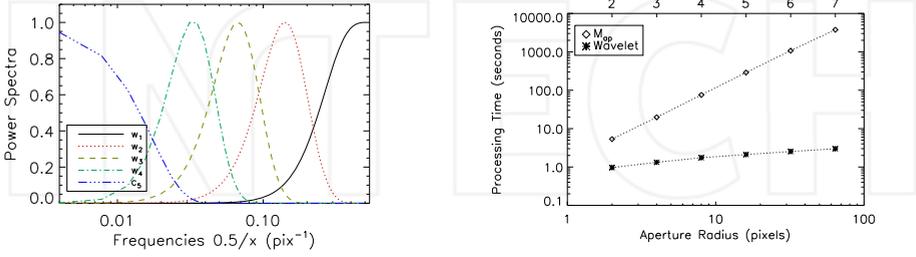
the shear catalogue directly. This avoids the need to generate an aperture mass map, which can be computationally intensive, and furthermore does not require the computation of the convergence,  $\kappa$ .

Computing the convergence from the shear measurements can be tricky for a number of reasons. Equation (4) describes a convolution over all 2D space. Given that we aim to invert this equation for an image of a finite size, a direct inversion in Fourier space will give rise to significant edge effects, and a leakage of power into so-called B-modes, which effectively imply a spurious (non-lensing) cross-component to the shear field  $\gamma$ , and usually only arise due to systematic effects in the lensing measurements.

Several methods have been devised to invert equation (4) whilst minimising these undesirable effects [26–29]. Most recently, the authors in [30] have presented a method based on a wavelet-Helmholtz decomposition, with which they demonstrate a reconstruction error at the few percent level, as compared to an error of  $\sim 30\%$  seen with Fourier-based methods. This implies that there is little advantage to the shear catalogue as opposed to the convergence map. Indeed, working on the convergence map and under the wavelet framework may offer some distinct advantages.

The first advantage comes directly from the nature of the wavelet transform. The aperture mass filter is a bandpass filter. Because the  $M_{ap}$  reconstruction is only computed for a discrete set of aperture scales, information on intermediate scales may be lost, particularly

at both the small and large frequency extremes. In contrast, the wavelet transform retains information on all scales. The first wavelet scale is effectively a high-pass filter, retaining all the high-frequency information in the image, while the remaining wavelet scales are bandpass filtered as with the  $M_{app}$ -filtered images. Finally, the wavelet transform retains  $c_J$ , which encodes the large-scale information in the image, and therefore consists of a smoothed version of the image (or an image with a low-pass filter applied). Figure 3(a) demonstrates this point for a wavelet transform with  $J = 5$ , showing the Fourier-space response of the wavelet filter as a function of scale  $j$ .



(a) The filter response of each of 5 scales of the wavelet transform

(b) Comparison of a the processing time of a brute-force aperture mass algorithm and the starlet transform algorithm

**Figure 3.** Illustration of the various advantages offered by the wavelet transform in terms of (a) retention of information on all scales and (b) algorithm processing time.

In addition, from a mapping perspective, the starlet wavelet transform algorithm is substantially faster than a brute-force computation of the aperture mass statistic in real space. Such a naive implementation has complexity  $\propto \mathcal{O}(N^2\theta_{cut}^2)$ , where  $N \times N$  is the dimension of the image, and  $\theta_{cut}$  is the chosen truncation radius. This scaling means that for large apertures or, equivalently, for high-resolution images, the algorithm may prove to be very time-consuming. The starlet wavelet transform algorithm is of complexity  $\propto \mathcal{O}(N^2J)$ , where  $J$  is the number of scales considered, and is limited by  $N \geq 2^J$ . This means that the processing time for the wavelet transform algorithm is sensitive only to the number of scales considered, rather than the size of the filter functions involved, and depends linearly on this number.

In Figure 3(b), we compare the processing time for the aperture mass algorithm and the starlet transform algorithm, both programmed in C++, to analyse an image of  $1024 \times 1024$  pixels on a  $2 \times 2.66$ GHz Intel Xeon Dual-Core processor. We consider aperture scales  $\theta = [2, 4, 8, 16, 32, 64]$ , which correspond to  $J = j_{max} + 1 = [2, 3, 4, 5, 6, 7]$  wavelet scales in the wavelet transform. In the aperture mass algorithm, the filters are truncated at a radius of  $\theta_{cut} = \theta$ . For the application of filters which necessitate truncation at a much larger radius, we expect the computation time to be roughly an order of magnitude longer. Even at the smallest aperture radius, the wavelet transform is  $\sim 5 \times$  faster than the aperture mass algorithm. At  $\theta = 64$  pixels, the wavelet transform is  $\sim 1200 \times$  faster than the aperture mass algorithm. Note that the wavelet transform for a given value of  $J$  simultaneously computes the wavelet transform at all scales  $2^j$ ,  $0 < j \leq J - 1$ , in addition to the smoothed continuum map  $c_J$ , whilst the aperture mass algorithm computes the transform at a single scale  $\theta$ . We note further that the computational time for the wavelet transform for  $J = 7$  wavelet scales is still a factor of  $\sim 2 \times$  less than the computational time for the aperture mass algorithm at  $\theta = 2$  pixels.

As discussed in [31], this time advantage further extends to higher-order statistics of the  $M_{ap}$ , which are typically related directly to  $n$ -point correlation functions of the shear. In recent years, tree codes have been employed to speed up computation of  $n$ -point correlation functions. Typical tree codes to compute  $n$ -point correlation functions are  $\mathcal{O}(N_{gal} \log(N_{gal}))$  [32] on a shear catalogue. For a single Euclid exposure of  $0.5 \text{ deg}^2$ , we can expect  $N_{gal} \sim 54,000$  (30 galaxies/arcmin<sup>2</sup>) - 180,000 (100 galaxies/arcmin<sup>2</sup>). Tree codes exist that act on pixelated data [33] which run at  $\mathcal{O}(N_{pix}^2 n_{bin})$  where  $N_{pix}^2$  is the total number of pixels and  $n_{bin}$  is the number of bins in the correlation function. For a Euclid exposure, assuming pixels of 1 arcminute, we have  $N_{pix}^2 = 1800$ , and  $n_{bin}$  will be dependent on the required resolution of the correlation function.

The wavelet method acts on pixelated data, and is  $\mathcal{O}(N^2)$  in computation time, so our algorithm will be comparable for computation of 2-point statistics, if the 2-point correlation function is computed on pixelated data, but if a shear catalogue is used, we will have a faster algorithm by at least an order of magnitude. For higher-order statistics, this advantage is even more pronounced. Furthermore, while optimised software is freely and publicly available to compute the wavelet transform, such optimised software is not available for  $n$ -point correlation functions.

Another advantage of the wavelet formalism is that it is possible to carry out an explicit denoising of the convergence field using thresholding based on a False Discovery Rate method. For details on this method, see [34], where the *MRLens* software package encoding this method is presented<sup>3</sup>. This allows one to derive robust detection levels in wavelet space, and to produce high-fidelity denoised mass maps.

In addition, wavelet-based methods offer more flexibility than aperture mass filters. Whilst we have thus far discussed only the starlet wavelet function, many other wavelet dictionaries may be used. The starlet filter seems ideal for lensing studies, due to its similarity to the aperture mass filter presented here, which was deemed to be optimal in [22]. However, different dictionaries may be optimal in different applications; for example, if one were attempting to study filamentary structure, ridgelets or curvelets might be a more appropriate basis. The vastness of the wavelet libraries and the public availability of fast algorithms to compute these transforms are major strengths of wavelet-based approaches.

#### 4. Weak lensing beyond two-point shear statistics

As noted previously, the optimality of second-order statistics to constrain cosmological parameters depends heavily on the assumption of Gaussianity of the field. However the weak lensing field is composed, at small scales, of non-Gaussian features such as clusters of galaxies. These non-Gaussian signatures carry additional information that cannot be extracted with second-order statistics. Many studies [35–39] have shown that combining second-order statistics with higher-order statistics tightens the constraints on cosmological parameters. We now consider the higher-order statistics most commonly used in weak lensing studies aimed at detecting and constraining non-Gaussianity in the lensing field.

<sup>3</sup> The *MRLens* software, along with many other software packages for astronomical applications, is freely available here: <http://cosmostat.org/software.html>

## 4.1. Higher-order lensing statistics

As we have already noted, it is more convenient to consider higher-order statistics of the  $M_{ap}$  or wavelet transform of the convergence field, as opposed to statistics of the shear field directly. The obvious first extension to two-point statistics is to consider the three-point correlation function or, equivalently, the bispectrum, usually considered as a function of aperture or wavelet scale.

Interesting analytical results relative to the shear three-point correlation function or the convergence bispectrum have been reported (for example, see [40–43]). However, it has been shown that in using only the equilateral configuration of the bispectrum, the ability to discriminate between cosmological models is relatively poor [38]. An analytical comparison has been performed in [39] between the full bispectrum and an optimal match-filter peak count for a Euclid-like survey, and both approaches were found to provide similar results. However, as the full bispectrum calculation has a much higher complexity than other statistics, and no public software exists to compute it, we will not consider the full bispectrum here. Rather, we will restrict ourselves to statistics that are more straightforward to compute from the  $M_{ap}$  or wavelet transform maps.

### 4.1.1. The Skewness

The skewness of the aperture mass map,  $\langle M_{ap}^3 \rangle$ , is the third-order moment of the aperture mass  $M_{ap}(\theta)$  and can be computed directly from shear maps filtered with different aperture mass. The skewness is a measure of the asymmetry of the probability distribution function. The probability distribution function will be more or less skewed positively depending on the abundance of dark matter haloes at the  $\theta$  scale. The formalism exists to predict the skewness of the aperture mass map for a given cosmological model, which is related to the three-point correlation function or the bispectrum of density fluctuations  $\delta$  [20, 37].

In [37], it is argued that the skewness as a function of scale is a preferable statistic to three-point correlation functions of the shear, as the integral relations between  $\langle M_{ap}^3 \rangle$  and the bispectrum are much easier and faster to compute than the three-point correlation function. This is because the skewness is a local measure of the bispectrum, whereas the integral kernel for the three-point correlation function is a highly oscillating function with infinite support.

This statistic can equivalently be computed from the wavelet transform of the convergence map as the skewness  $\langle w_j^3 \rangle$  of the wavelet band  $j$  corresponding to the aperture scale  $\theta$ . This can be computed either on the noisy convergence map estimated, for example, by the Fourier space relationship between the shear and convergence – in which case the results should be comparable to the  $M_{ap}$  results – or on a denoised convergence map using the method of [34], which should provide improved results.

### 4.1.2. The Kurtosis

The kurtosis of the aperture mass map,  $\langle M_{ap}^4 \rangle$ , is the fourth-order moment of the aperture mass  $M_{ap}(\theta)$  and can be computed directly from the different aperture mass maps. The kurtosis is a measure of the peakedness of the probability distribution function. The presence of dark matter haloes at a given  $\theta$  scale will flatten the probability distribution function and widen its shoulders leading to a larger kurtosis. The formalism exists to predict the kurtosis

of the aperture mass map for a given cosmological model, which is related to the four-point correlation function or trispectrum of the 3D density contrast [20].

Again, this statistic can be computed from the wavelet transform of the convergence map as the kurtosis  $\langle w_j^4 \rangle$  of the wavelet band  $j$  corresponding to the aperture scale  $\theta$ . This can be computed either on a noisy or a denoised convergence map, the latter expected to yield improved constraints.

#### 4.1.3. Peak Counts

We define a peak as set of connected pixels above a detection threshold  $\mathcal{T}$ , and denote the peak count in the  $M_{ap}$  and wavelet maps by  $P_{M_{ap}}^{\mathcal{T}}$  and  $P_{w_j}^{\mathcal{T}}$ , respectively. If peak counting is carried out on a denoised map, the detection threshold  $\mathcal{T}$  is automatically set by the denoising algorithm, and a small threshold  $\epsilon$  is used when identifying peaks in the denoised maps, in order to reject spurious detections in these denoised maps.

We consider all pixels that are connected via the sides or the corners of a pixel as one structure. In a two-dimensional projected map of the convergence, we are therefore unable to discriminate between peaks due to massive halos and peaks due to projections of large-scale structures such as filaments.

While theory exists to predict cluster counts from the halo model and a cosmological model encoding the growth and evolution of structures in the Universe, there is no analytic formalism to predict the fraction of spurious detections in lensing maps arising from projections of large-scale structure. In [44], an attempt is made to derive an analytic formalism for predicting peak counts in projected maps. However, this method is based on an assumption of Gaussianity in the lensing field and, predictably, underestimates counts at the high end of the mass function, where clusters of galaxies are dominant. This means that predicting peak counts for a given cosmological model relies on considering N-body simulations generated under a range of cosmological parameters (for example, see [45]).

## 4.2. Optimal capture of non-Gaussianity

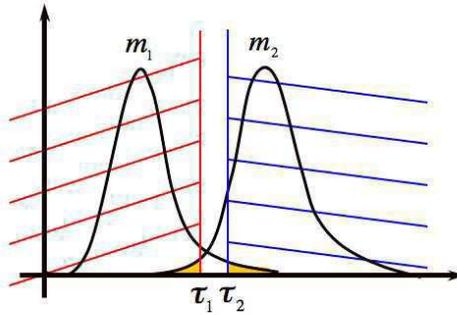
The question now arises: which of these statistics provides the most information about the underlying cosmology and – specifically – non-Gaussianity in the density field? In order to assess the performance of these statistics, we consider the ability of each statistic to discriminate between cosmological models using N-body simulations under a range of different cosmologies [38, 46].

To this end, we consider N-body simulations carried out for 5 different cosmological models along the  $\Omega_M - \sigma_8$  degeneracy, the parameters of which are summarised in Table 1 below [46]. These simulations were carried out using the RAMSES N-body code [47], and full details of the models considered are given in [38]. For each model, 100 realizations of weak lensing maps were generated for a field of  $3.95^\circ \times 3.95^\circ$ , downsampled to  $1024 \times 1024$  pixels ( $0.23'$  per pixel). Noise was added to the simulations at a level consistent with predictions for deep space-based observations, with a number density of galaxies of  $n_g = 100 \text{ gal/arcmin}^2$ . This is somewhat optimistic; however, as seen in [46], increasing the noise level in the simulations does not change the conclusions of the comparison test between the higher-order statistics.

Model	Box Length ( $h^{-1}\text{Mpc}$ )	$\Omega_M$	$\Omega_\Lambda$	$h$	$\sigma_8$
$m_1$	165.8	0.23	0.77	0.594	1
$m_2$	159.5	0.3	0.7	0.7	0.9
$m_3$	152.8	0.36	0.64	0.798	0.8
$m_4$	145.7	0.47	0.53	0.894	0.7
$m_5$	137.5	0.64	0.36	0.982	0.6

**Table 1.** Parameters of the five cosmological models that have been chosen along the  $\Omega_M - \sigma_8$  degeneracy. The simulations have  $256^3$  particles,  $h$  is equal to  $H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , where  $H_0$  is the Hubble constant.

To find the best statistic, we need to characterise quantitatively for each statistic the discrimination between two different models  $m_1$  and  $m_2$ . To do this, we consider the distribution functions for the different statistics estimated on the 100 realisations of each model. These distributions are expected to overlap, so False Discovery Rate (FDR) method is used to determine the threshold  $\tau_1$  in the distribution function of model  $m_1$ , such that the distribution function of model  $m_2$  accounts for fewer than a fraction  $\alpha = 0.05$  of the total counts. A similar threshold  $\tau_2$  is defined for the distribution function of model  $m_2$ . This is illustrated in Figure 4. The discrimination efficiency is then defined as the percentage of counts remaining under the hatched area for each model.



**Figure 4.** Illustration of the discrimination efficiency criterion.  $\tau_1$  and  $\tau_2$  are defined such that the yellow area under the curve represents 0.05 of the total area under the curves for models  $m_2$  and  $m_1$ , respectively. The discrimination efficiency is defined as the area under the  $m_i$  curve delimited by the threshold  $\tau_i$ , as a percentage of the total area under the model  $m_i$  curve.

In Table 2 below, we show the mean discrimination efficiency obtained considering all pairs of models for the higher-order statistics of interest. These are presented as a function of  $M_{ap}$  or wavelet scale, and shown for statistics computed on the aperture mass map and the wavelet transform of the noisy convergence map. The results overall are comparable, with the wavelet transform statistics appearing to offer a slight improvement over the  $M_{ap}$  statistics in all cases. This is perhaps related to the fact that, when computing the  $M_{ap}$  statistics, we truncate the filter at  $\vartheta_{cut} = \theta$ , which may give rise to some small systematics in the resulting maps. Furthermore, it is clear that, in all cases, the peak statistics are much more efficient at discriminating between cosmological models than either the skewness or the kurtosis.

$\theta_i$	$\langle M_{ap}^3 \rangle$	$\langle M_{ap}^4 \rangle$	$P_{M_{ap}}^{2\sigma}$	$P_{M_{ap}}^{3\sigma}$
0.46'	04.60 %	02.30 %	39.70 %	54.95 %
0.92'	<b>33.40 %</b>	03.70 %	79.30 %	76.40 %
1.85'	03.45 %	01.45 %	<b>91.25 %</b>	<b>89.20 %</b>
3.70'	15.15 %	23.00 %	69.40 %	86.70 %
7.40'	26.95 %	<b>24.30 %</b>	4.90 %	60.50 %

(a) Mean discrimination efficiency of  $M_{ap}$  statistics

Scale	$\langle w_j^3 \rangle$	$\langle w_j^4 \rangle$	$P_{w_j}^{2\sigma}$	$P_{w_j}^{3\sigma}$
0.46'	02.00 %	01.15 %	12.05%	00.70 %
0.92'	<b>37.95 %</b>	04.75 %	86.30 %	73.05 %
1.85'	03.55 %	02.10 %	<b>94.40 %</b>	<b>93.85 %</b>
3.70'	18.25 %	25.65 %	84.05 %	87.05 %
7.40'	36.40 %	<b>30.90 %</b>	24.60 %	66.35 %

(b) Mean discrimination efficiency of wavelet statistics

**Table 2.** Mean discrimination efficiency (in percent) from noisy  $M_{ap}$  reconstructions of the shear field, compared with that obtained on the wavelet transform of the noisy convergence maps.

In Table 3 below, we present the mean discrimination efficiency of statistics computed on the denoised convergence maps, as well as the peak discrimination efficiency between each pair of models in our sample. There is a clear improvement on the discrimination efficiency of all the higher-order statistics when denoising is applied to the convergence maps. This is unsurprising, as the presence of Gaussian (or near-Gaussian) noise within the shear and convergence maps will make the whole field appear more Gaussian, masking the non-Gaussian features in the map and pushing the skewness and kurtosis values closer to zero. The more noise that is present in the maps, the stronger this effect will be. It is for this reason that such a dramatic improvement is seen in these statistics when denoising is applied to the convergence maps before the statistics are computed.

The improvement seen in the discrimination efficiency of the peak statistics is also significant, and this arises due to the fact that while the peak counts in the noisy maps involve application of a simple  $n\sigma$  detection threshold to the maps, the denoising algorithm involves a more sophisticated discriminant between signal and noise, again making use of the FDR method. This technique is more effective and distinguishing between signal and noise, and therefore more information about the true peaks in the map is retained using this denoising method as compared to the  $n\sigma$  thresholding applied previously.

It is clear that peak statistics are much more efficient at capturing non-Gaussianity in the shear field than other higher-order statistics, and that it is advantageous to measure this statistic on denoised convergence maps. However, in order to combine constraints from peak counting with other probes, such as two-point statistics, in order to better constrain the cosmological model, it is important to be able to predict peak statistics as a function of cosmology in order to be able to compute the likelihood function for all models within the parameter space. As we have discussed, this is not possible to do analytically, due to contamination by projections of large scale structures, and it is too computationally-intensive

Scale	$\langle \tilde{w}_j^3 \rangle$	$\langle \tilde{w}_j^4 \rangle$	$P_{\tilde{w}_j}$
0.46'	53.40 %	43.20 %	68.35 %
0.92'	47.90 %	41.15 %	92.45 %
1.85'	58.80 %	44.70 %	<b>96.75 %</b>
3.70'	<b>63.30 %</b>	<b>48.05 %</b>	90.40 %
7.40'	54.90 %	40.45 %	63.45 %

(a) Mean discrimination efficiency

	$m_1$	$m_2$	$m_3$	$m_4$	$m_5$
$m_1$	x	85 %	100 %	100 %	100 %
$m_2$	89 %	x	92 %	100 %	100 %
$m_3$	100 %	92 %	x	89 %	100 %
$m_4$	100 %	100 %	92 %	x	98 %
$m_5$	100 %	100 %	100 %	98 %	x

(b) Discrimination efficiency for all models obtained using peak statistics

**Table 3.** Discrimination efficiency for statistics computed on denoised convergence maps.

to consider obtaining the predictions, and associated covariances, from N-body simulations if we wish to sample the parameter space completely and at high resolution.

As the quality of data available to astronomers continues to improve, it has recently become possible to consider the shear field in three dimensions, using the colours of galaxies to determine their redshifts and, therefore, distances from us. If this information may be used to deproject the lensing signal, and thus recover information about the full three-dimensional density field, it should then be possible offer predictions for peak statistics as a function of cosmology, uncontaminated by projection effects which will, in turn, allow us to improve our constraints on our cosmological model.

## 5. Reconstructing the density contrast in 3D

The measured shear can be related to the 2D convergence via a simple linear relationship (equation (4)), and the convergence is related to the density contrast by another linear mapping (equation (6)). We can express these relationships conveniently in matrix notation as

$$\kappa(\boldsymbol{\theta}, z) = \mathbf{Q}\delta(\boldsymbol{\theta}, z), \quad (21)$$

$$\gamma(\boldsymbol{\theta}, z) = \mathbf{P}_{\gamma\kappa}\mathbf{Q}\delta(\boldsymbol{\theta}, z). \quad (22)$$

The 3D lensing problem is therefore one of finding an estimator to invert equation (21) or (22) in the presence of noise. For simplicity, in what follows we assume the shear measurement noise is Gaussian and uncorrelated between redshift bins. In practice, errors in the photometric estimates of the redshifts of galaxies will introduce correlations between redshift bins. Methods have been developed to account for such errors, however, and the problem is therefore readily tractable [48].

Therefore the 3D lensing problem is effectively one of observing the density contrast convolved with the linear operator  $\mathbf{R}$ , and contaminated by noise, which is assumed to be Gaussian. Formally, we can write

$$\mathbf{d} = \mathbf{R}\mathbf{s} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \quad (23)$$

where  $\mathbf{d}$  is the observation,  $\mathbf{s}$  the real density and  $\varepsilon$  the Gaussian noise.

### 5.1. Linear approaches to 3D map-making

The general idea behind linear inversion methods is to find a linear operator  $\mathbf{H}$  that acts on the data vector to yield a solution that minimises some functional, such as the variance of the residual between the estimated signal and the true signal, subject to some regularisation or prior-based constraints. Two different linear approaches have been described in recent literature [4, 49].

The most competitive method is that proposed in [4], in which the authors propose a Saskatoon filter [50, 51], which combines a Wiener filter and an inverse variance filter, with a tuning parameter  $\alpha$  introduced that allows switching between the two. This gives rise to a minimum variance filter, expressed as

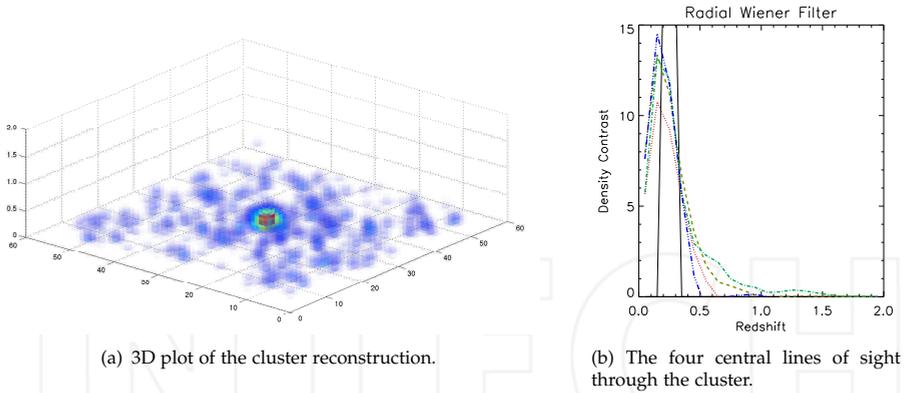
$$\hat{\mathbf{s}}_{MV} = [\alpha \mathbf{1} + \mathbf{S}\mathbf{R}^\dagger \mathbf{\Sigma}^{-1} \mathbf{R}]^{-1} \mathbf{S}\mathbf{R}^\dagger \mathbf{\Sigma}^{-1} \mathbf{d}, \quad (24)$$

where  $\mathbf{S} \equiv \langle \mathbf{s}\mathbf{s}^\dagger \rangle$  encodes prior information about the signal covariance,  $\mathbf{\Sigma} \equiv \langle \mathbf{n}\mathbf{n}^\dagger \rangle$  gives the covariance matrix of the noise, and  $\mathbf{1}$  is the identity matrix.

This switching is designed to allow a balance between the increased constraining power offered by the Wiener filter over the inverse variance filter – which yields an improved signal-to-noise in the reconstruction – and the biasing that the Wiener filter imposes on the solution.

As discussed extensively in [4] and [49], linear methods give rise to a significant bias in the location of detected peaks, damping of the peak signal, and a substantial smearing of the density along the line of sight. Furthermore, the resolution attainable in the reconstructions obtained using linear methods is limited to the resolution of the data. In other words, we cannot reconstruct the density contrast field at higher resolution than the resolution of our input data which, in turn, is limited by the noise properties of the data.

Figure 5 shows the 3D reconstruction obtained in this way for a simulated cluster of galaxies at a redshift of  $z_{cl} = 0.25$ . The tuning parameter used for this reconstruction was  $\alpha = 0.05$ . The cluster was simulated according to an NFW halo with  $M = 10^{15} h^{-1} M_\odot$  and  $c = 3$ , and the shear data were assumed to come from a galaxy distribution given by  $p(z) \propto z^2 e^{-(1.4z)^{1.5}}$  [52, 53], with a maximum redshift of  $z_{max} = 2.0$  and a galaxy density of  $n_g = 100$  galaxies/arcmin<sup>2</sup>. The simulation covers a  $1^\circ \times 1^\circ$  field binned into  $60 \times 60$  angular pixels, and 20 redshift bins. Shown in the figure are both the 3D rendering of the reconstruction, and a 1D plot showing the four central lines of sight through the cluster as a function of redshift. The smearing, damping and redshift bias effects are all clearly visible in the 1D plot, where the amplitude of the cluster density contrast should be  $\delta \sim 36$ .



**Figure 5.** The reconstruction of a simulated cluster of galaxies using the method of [4]. Coloured broken lines in the 1D plot show the reconstruction of the four central lines of sight, while the solid line shows the true cluster density contrast.

## 5.2. Compressive sensing

We consider some data  $Y_i$  ( $i \in [1, \dots, m]$ ) acquired through the linear system

$$Y = \Theta X, \quad (25)$$

where  $\Theta$  is an  $m \times n$  matrix. Compressed sensing [54, 55] is a sampling/compression theory based on the sparsity of the observed signal, which shows that, under certain conditions, one can exactly recover a  $k$ -sparse signal (a signal for which only  $k$  pixels have values different from zero, out of  $n$  total pixels, where  $k < n$ ) from  $m < n$  measurements.

This recovery is possible from undersampled data only if the sensing matrix  $\Theta$  verifies the *restricted isometry property* (RIP) [see 54, for more details]. This property has the effect that each measurement  $Y_i$  contains some information about the entire pixels of  $X$ ; in other words, the sensing operator  $\Theta$  acts to spread the information contained in  $X$  across many measurements  $Y_i$ .

Under these two constraints – sparsity and a transformation meeting the RIP criterion – a signal can be recovered exactly even if the number of measurements  $m$  is much smaller than the number of unknown  $n$ . This means that, using CS methods, we will be able to outperform the well-known Shannon sampling criterion by far.

The solution  $X$  of (25) is obtained by minimizing

$$\min_X \|X\|_1 \quad s.t. \quad Y = \Theta X, \quad (26)$$

where the  $\ell_1$  norm is defined by  $\|X\|_1 = \sum_i |X_i|$ . The  $\ell_1$  norm is well-known to be a sparsity-promoting function; i.e. minimisation of the  $\ell_1$  norm yields the most sparse solution to the inverse problem. Many optimisation methods have been proposed in recent years to minimise this equation. More details about CS and  $\ell_1$  minimisation algorithms can be found in [6].

In real life, signals are generally not “strictly” sparse, but are *compressible*; i.e. we can represent the signal in a basis or frame (Fourier, Wavelets, Curvelets, etc.) in which the curve obtained by plotting the obtained coefficients, sorted by their decreasing absolute values, exhibits a polynomial decay. Note that most natural signals and images are compressible in an appropriate basis.

We can therefore reformulate the CS equation above (Equation (26)) to include the data transformation matrix  $\Phi$ :

$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t. } Y = \Theta\Phi\alpha, \quad (27)$$

where  $X = \Phi^*\alpha$ , and  $\alpha$  are the coefficients of the transformed solution  $X$  in  $\Phi$ , which is generally referred to as the *dictionary*. Each column represents a vector (also called an *atom*), which ideally should be chosen to match the features contained in  $X$ . If  $\Phi$  admits a fast implicit transform (e.g. Fourier transform, Wavelet transform), fast algorithms exist to minimise Equation (27).

One problem we face when considering CS in a given application is that very few matrices meet the RIP criterion. However, it has been shown that accurate recovery can be obtained as long the mutual coherence between  $\Theta$  and  $\Phi$ ,  $\mu_{\Theta,\Phi} = \max_{i,k} |\langle \Theta_i, \Phi_k \rangle|$ , is low [56]. The mutual coherence measures the degree of similarity between the sparsifying basis and the sensing operator. Hence, in its relaxed definition, we consider a linear inverse problem  $Y = \Theta\Phi X$  as being an instance of CS when

1. the problem is underdetermined,
2. the signal is compressible in a given dictionary  $\Phi$ ,
3. the mutual coherence  $\mu_{\Theta,\Phi}$  is low. This will happen every time the matrix  $\mathbf{A} = \Theta\Phi$  has the effect of spreading out the coefficients  $\alpha_j$  of the sparse signal on all measurements  $Y_i$ .

Most CS applications described in the literature are based on such a soft CS definition. Compressed sensing was introduced for the first time in astronomy for data compression [57, 58], and a direct link between CS and radio-interferometric image reconstruction was recently established in [59], leading to dramatic improvement thanks to the sparse  $\ell_1$  recovery [60].

The 3D weak lensing reconstruction problem can be seen to completely meet the soft-CS criteria above. The problem is underdetermined, as we seek a higher resolution than can be attained in the noise-limited observations, the matter density in the Universe is sparsely distributed, and the lensing operator spreads out the underlying density in a compressed sensing way.

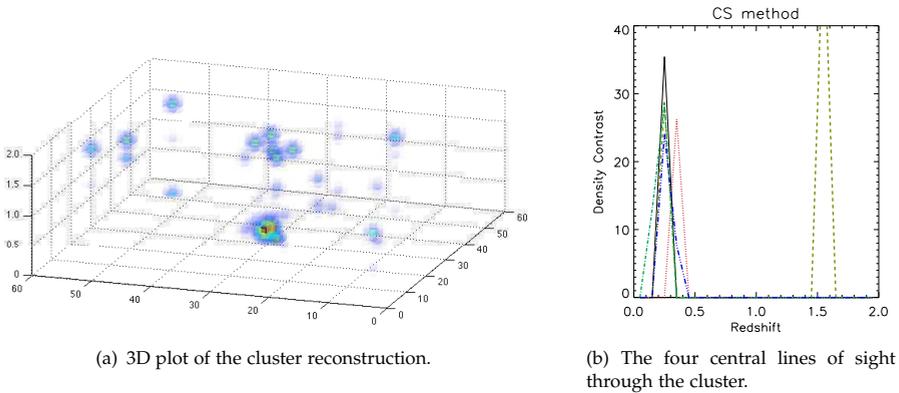
In particular, for the reconstruction of clusters of galaxies, we are in a perfect situation for sparse recovery because clusters are localised in the angular domain, and are not resolved along the line of sight owing to the bin size. They can therefore be modelled as Dirac  $\delta$ -functions along the line of sight, while an isotropic wavelet basis can be used in the angular domain.

### 5.3. Results and future prospects

In [61], we present an algorithm to solve the 3D lensing problem. In this method, the 3D lensing problem is reduced to a one-dimensional problem, by taking as the data vector the (noisy) lensing convergence along each line of sight, which is related to the density contrast through Equation (21). Each line of sight can therefore be considered independently.

Clearly, a one-dimensional implementation throws away information, because we do not account at all for the correlation between neighbouring lines of sight that will arise in the presence of a large structure in the image; however, reducing the problem to a single dimension is fast and easy to implement, and allows us to test the efficacy of the algorithm using a particularly simple basis function through which we impose sparsity.

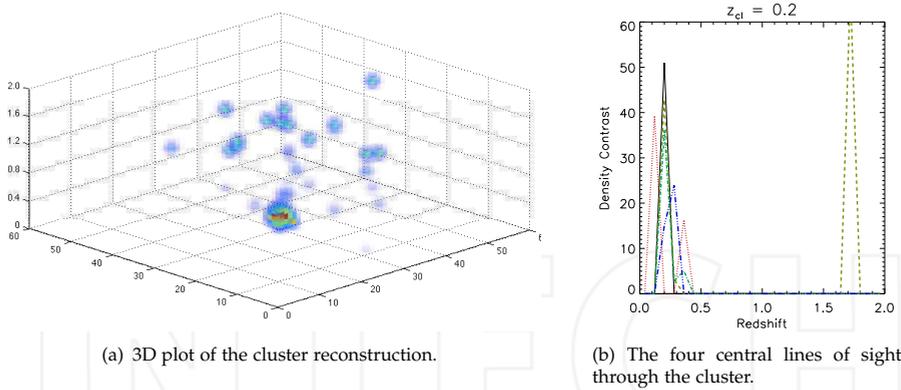
In Figure 6, we show the reconstruction obtained in this way for the simulated cluster described in section 5.1. The line of sight plot shows a clear improvement in all the target areas, reducing the bias, smearing and damping effects seen using linear methods. Small-scale noise is present, particularly at high redshifts, but this is likely due to overfitting in the algorithm used, and may be reduced by considering the whole 3D field, rather than each line of sight independently (for a full discussion of this issue, see [61]).



**Figure 6.** The reconstruction of a simulated cluster of galaxies using the method of [61]. Coloured broken lines in the 1D plot show the reconstruction of the four central lines of sight, while the solid line shows the true cluster density contrast.

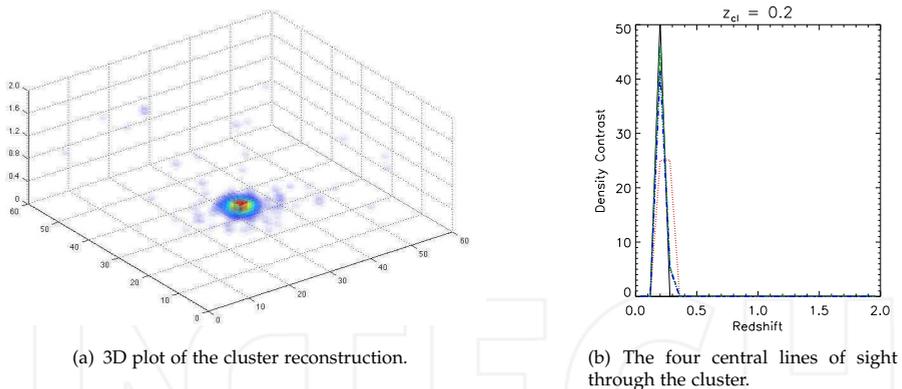
This reconstruction was undertaken using the same resolution on the reconstruction as on the input data. However, we know that the CS approach is particularly well-adapted to dealing with ill-posed inversion problems. In order to test this, we consider a cluster at a redshifts of  $z_{cl} = 0.2$ , simulated as before. We use  $N_{sp} = 20$  redshift bins in our data, but now aim to reconstruct our density contrast with a redshift resolution of  $N_{lp} = 25$ . The results are shown in Figure 7. Here, again, we see small scale noise at high-redshift, but the overall smearing and redshift bias issues seen in the linear reconstructions is absent.

As noted previously, the one-dimensional solver employed here throws away a wealth of information about the angular correlation of the lensing signal, and is thus not optimal. Indeed, a simple algorithm based on this CS approach, but implemented as a full three-dimensional treatment, offers marked improvements in the quality of the reconstructions [62].



**Figure 7.** The reconstruction of a simulated cluster of galaxies at a redshift of  $z_{cl} = 0.2$ , and with improved resolution on the reconstruction as compared to the input data.

This is demonstrated in Figure 8, which shows reconstructions of the cluster at redshift  $z_{cl} = 0.2$ , with the same improved resolution in the reconstruction as before, but this time using the three-dimensional CS approach. Dramatic improvement is seen on all fronts, with the reconstructions showing no bias or redshift smearing, and very little amplitude damping, and with none of the small-scale false detections seen in the 1D CS approach.



**Figure 8.** The reconstruction of a simulated cluster of galaxies at a redshift of  $z_{cl} = 0.2$  using a 3D CS approach, and with improved resolution on the reconstruction as compared to the input data.

This marked improvement in the cluster reconstruction seen in the 3D CS approach represents a definite step in the right direction for weak lensing studies. There is much work to be done, however. The application of this CS approach to the 3D lensing problem is a very recent development, and many questions remain: the choice of algorithm, for example; how best to control the noise in the reconstruction; how to deal with photometric redshift errors; by what factor the reconstruction resolution might be improved as compared to the data.

Yet it is clear that this approach opens up the possibility of being able to generate accurate reconstructions of the density contrast using weak lensing measurements, and perhaps using information such as the 3D peak count – in combination with constraints from other probes – to place ever-tighter constraints on our cosmological model. This, in turn, will offer a unique insight into the nature and properties of the dark components of the Universe.

## Acknowledgements

This work is supported by the European Research Council grant SparseAstro (ERC-228261). The authors would like to thank Jalal Fadili for useful discussions, and Jake VanderPlas for helpful comments and use of his code to generate the linear reconstructions.

## Author details

Adrienne Leonard<sup>1,\*</sup>, Jean-Luc Starck<sup>1</sup>,  
Sandrine Pires<sup>1</sup> and François-Xavier Dupé<sup>2</sup>

\* Address all correspondence to: adrienne.leonard@cea.fr

1 Laboratoire AIM (UMR 7158), CEA/DSM-CNRS-Université Paris Diderot, IRFU, SEDI-SAP, Service d'Astrophysique, Centre de Saclay, France

2 Laboratoire d'Informatique Fondamentale, Centre de Mathématique et d'Informatique, UMR CNRS, Aix-Marseille Université, Marseille, France

## References

- [1] D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters. *The Astrophysical Journal Supplement*, 192:16–+, 2011.
- [2] G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, N. Jarosik, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data Processing, Sky Maps, and Basic Results. *The Astrophysical Journal Supplement*, 180:225–245, 2009.
- [3] J. A. Frieman, M. S. Turner, and D. Huterer. Dark Energy and the Accelerating Universe. *Annual Reviews in Astronomy and Astrophysics*, 46:385–432, 2008.
- [4] P. Simon, A. N. Taylor, and J. Hartlap. Unfolding the matter distribution using three-dimensional weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 399:48–68, 2009.
- [5] M.J. Fadili and J.-L. Starck. Monotone operator splitting for optimization problems in sparse recovery. In *Proceedings of the International Conference on Image Processing, ICIP 2009, 7-10 November 2009, Cairo, Egypt*, pages 1461–1464. IEEE, 2009.

- [6] J.-L. Starck, F. Murtagh, and M.J. Fadili. *Sparse Image and Signal Processing*. Cambridge University Press, 2010.
- [7] A. Refregier, A. Amara, T. D. Kitching, A. Rassat, R. Scaramella, J. Weller, and f. t. Euclid Imaging Consortium. Euclid Imaging Consortium Science Book. *ArXiv:1001.0061*, 2010.
- [8] P. Schneider. *Weak Gravitational Lensing*, pages 269–+. Springer, 2006.
- [9] M. Bartelmann and P. Schneider. Weak gravitational lensing. *Physics Reports*, 340:291–472, 2001.
- [10] P. Schneider, L. van Waerbeke, M. Kilbinger, and Y. Mellier. Analysis of two-point statistics of cosmic shear. I. Estimators and covariances. *Astronomy and Astrophysics*, 396:1–19, 2002.
- [11] N. Kaiser. Weak Lensing and Cosmology. *The Astrophysical Journal*, 498:26, 1998.
- [12] E. Semboloni, Y. Mellier, L. van Waerbeke, H. Hoekstra, I. Tereno, K. Benabed, S. D. J. Gwyn, L. Fu, M. J. Hudson, R. Maoli, and L. C. Parker. Cosmic shear analysis with CFHTLS deep data. *Astronomy and Astrophysics*, 452:51–61, 2006.
- [13] T. Schrabback, J. Hartlap, B. Joachimi, M. Kilbinger, P. Simon, K. Benabed, M. Bradač, T. Eifler, T. Erben, C. D. Fassnacht, F. W. High, S. Hilbert, H. Hildebrandt, H. Hoekstra, K. Kuijken, P. J. Marshall, Y. Mellier, E. Morganson, P. Schneider, E. Semboloni, L. van Waerbeke, and M. Velander. Evidence of the accelerated expansion of the Universe from weak lensing tomography with COSMOS. *Astronomy and Astrophysics*, 516:A63, 2010.
- [14] P. Schneider and M. Lombardi. The three-point correlation function of cosmic shear. I. The natural components. *Astronomy and Astrophysics*, 397:809–818, 2003.
- [15] M. Takada and B. Jain. The Three-Point Correlation Function for Spin-2 Fields. *The Astrophysical Journal Letters*, 583:L49–L52, 2003.
- [16] M. Zaldarriaga and R. Scoccimarro. Higher Order Moments of the Cosmic Shear and Other Spin-2 Fields. *The Astrophysical Journal*, 584:559–565, 2003.
- [17] P. Schneider. Detection of (dark) matter concentrations via weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 283:837–853, 1996.
- [18] P. Schneider, L. van Waerbeke, B. Jain, and G. Kruse. A new measure for cosmic shear. *Monthly Notices of the Royal Astronomical Society*, 296:873–892, 1998.
- [19] L. van Waerbeke. Scale dependence of the bias investigated by weak lensing. *Astronomy and Astrophysics*, 334:1–10, 1998.
- [20] M. Jarvis, G. Bernstein, and B. Jain. The skewness of the aperture mass statistic. *Monthly Notices of the Royal Astronomical Society*, 352:338–352, 2004.
- [21] R. G. Crittenden, P. Natarajan, U.-L. Pen, and T. Theuns. Discriminating Weak Lensing from Intrinsic Spin Correlations Using the Curl-Gradient Decomposition. *The Astrophysical Journal*, 568:20–27, 2002.

- [22] T.-J. Zhang, U.-L. Pen, P. Zhang, and J. Dubinski. Optimal Weak-Lensing Skewness Measurements. *The Astrophysical Journal*, 598:818–826, 2003.
- [23] J.-L. Starck, F. D. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis*. Cambridge University Press, 1998.
- [24] J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer, 2006. 2nd edn.
- [25] J.-L. Starck, F. Murtagh, and M. Bertero. in *Handbook of Mathematical Methods in Imaging*. Springer, 2011. in press.
- [26] S. Seitz and P. Schneider. A new finite-field mass reconstruction algorithm. *Astronomy and Astrophysics*, 374:740–745, 2001.
- [27] C. Seitz and P. Schneider. Steps towards nonlinear cluster inversion through gravitational distortions. III. Including a redshift distribution of the sources. *Astronomy and Astrophysics*, 318:687–699, 1997.
- [28] S. Seitz and P. Schneider. Cluster lens reconstruction using only observed local data: an improved finite-field inversion technique. *Astronomy and Astrophysics*, 305:383–+, 1996.
- [29] C. Seitz and P. Schneider. Steps towards nonlinear cluster inversion through gravitational distortions II. Generalization of the Kaiser and Squires method. *Astronomy and Astrophysics*, 297:287–+, 1995.
- [30] E. Deriaz, J.-L. Starck, and S. Pires. Wavelet Helmholtz decomposition for weak lensing mass map reconstruction. *Astronomy and Astrophysics*, 540:A34, 2012.
- [31] A. Leonard, S. Pires, and J.-L. Starck. Fast calculation of the weak lensing aperture mass statistic. *Monthly Notices of the Royal Astronomical Society*, 423:3405–3412, 2012.
- [32] Lucy Liuxuan Zhang and Ue-Li Pen. Fast n-point correlation functions and three-point lensing application. *New Astronomy*, 10(7):569 – 590, 2005.
- [33] H. K. Eriksen, P. B. Lilje, A. J. Banday, and K. M. Gęłski. Estimating n-point correlation functions from pixelized sky maps. *The Astrophysical Journal Supplement Series*, 151(1):1, 2004.
- [34] J.-L. Starck, S. Pires, and A. Réfrégier. Weak lensing mass reconstruction using wavelets. *Astronomy and Astrophysics*, 451:1139–1150, 2006.
- [35] F. Bernardeau, L. van Waerbeke, and Y. Mellier. Weak lensing statistics as a probe of  $\{\Omega_{\text{M}}\}$  and power spectrum. *Astronomy and Astrophysics*, 322:1–18, 1997.
- [36] M. Takada and B. Jain. Cosmological parameters from lensing power spectrum and bispectrum tomography. *Monthly Notices of the Royal Astronomical Society*, 348:897–915, 2004.

- [37] M. Kilbinger and P. Schneider. Cosmological parameters from combined second- and third-order aperture mass statistics of cosmic shear. *Astronomy and Astrophysics*, 442:69–83, 2005.
- [38] S. Pires, J.-L. Starck, A. Amara, A. Réfrégier, and R. Teyssier. Cosmological model discrimination with weak lensing. *Astronomy and Astrophysics*, 505:969–979, 2009.
- [39] J. Bergé, A. Amara, and A. Réfrégier. Optimal Capture of Non-Gaussianity in Weak-Lensing Surveys: Power Spectrum, Bispectrum, and Halo Counts. *The Astrophysical Journal*, 712:992–1002, 2010.
- [40] C.-P. Ma and J. N. Fry. Deriving the Nonlinear Cosmological Power Spectrum and Bispectrum from Analytic Dark Matter Halo Profiles and Mass Functions. *The Astrophysical Journal*, 543:503–513, 2000.
- [41] C.-P. Ma and J. N. Fry. Halo Profiles and the Nonlinear Two- and Three-Point Correlation Functions of Cosmological Mass Density. *The Astrophysical Journal Letters*, 531:L87–L90, 2000.
- [42] R. Scoccimarro and H. M. P. Couchman. A fitting formula for the non-linear evolution of the bispectrum. *Monthly Notices of the Royal Astronomical Society*, 325:1312–1316, 2001.
- [43] A. Cooray and W. Hu. Weak Gravitational Lensing Bispectrum. *The Astrophysical Journal*, 548:7–18, 2001.
- [44] M. Maturi, C. Angrick, F. Pace, and M. Bartelmann. An analytic approach to number counts of weak-lensing peak detections. *Astronomy and Astrophysics*, 519:A23, 2010.
- [45] J. P. Dietrich and J. Hartlap. Cosmology with the shear-peak statistics. *Monthly Notices of the Royal Astronomical Society*, 402:1049–1058, 2010.
- [46] S. Pires, A. Leonard, and J.-L. Starck. Cosmological constraints from the capture of non-Gaussianity in weak lensing data. *Monthly Notices of the Royal Astronomical Society*, 423:983–992, 2012.
- [47] R. Teyssier. Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. *Astronomy and Astrophysics*, 385:337–364, 2002.
- [48] Z. Ma, W. Hu, and D. Huterer. Effects of Photometric Redshift Uncertainties on Weak-Lensing Tomography. *The Astrophysical Journal*, 636:21–29, 2006.
- [49] J. T. VanderPlas, A. J. Connolly, B. Jain, and M. Jarvis. Three-dimensional Reconstruction of the Density Field: An SVD Approach to Weak-lensing Tomography. *The Astrophysical Journal*, 727:118–+, 2011.
- [50] M. Tegmark. How to Make Maps from Cosmic Microwave Background Data without Losing Information. *The Astrophysical Journal Letters*, 480:L87+, 1997.

- [51] M. Tegmark, A. de Oliveira-Costa, M. J. Devlin, C. B. Netterfields, L. Page, and E. J. Wollack. A High-Resolution Map of the Cosmic Microwave Background around the North Celestial Pole. *The Astrophysical Journal Letters*, 474:L77+, 1997.
- [52] A. N. Taylor, T. D. Kitching, D. J. Bacon, and A. F. Heavens. Probing dark energy with the shear-ratio geometric test. *Monthly Notices of the Royal Astronomical Society*, 374:1377–1403, 2007.
- [53] T. D. Kitching, A. F. Heavens, and L. Miller. 3D photometric cosmic shear. *Monthly Notices of the Royal Astronomical Society*, pages 426–+, 2011.
- [54] E.J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12), 2006. 5406–5425.
- [55] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 2006. 1289–1306.
- [56] E. J. Candes and Y. Plan. A probabilistic and ripples theory of compressed sensing. submitted, 2010.
- [57] J. Bobin, J.-L. Starck, and R. Ottensamer. Compressed Sensing in Astronomy. *IEEE Journal of Selected Topics in Signal Processing*, 2:718–726, 2008.
- [58] N. Barbey, M. Sauvage, J.-L. Starck, R. Ottensamer, and P. Chaniel. Feasibility and performances of compressed sensing and sparse map-making with Herschel/PACS data. *Astronomy and Astrophysics*, 527:A102+, 2011.
- [59] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, and P. Vandergheynst. Compressed sensing imaging techniques for radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 395:1733–1742, 2009.
- [60] F. Li, T. J. Cornwell, and F. de Hoog. The application of compressive sampling to radio astronomy. I. Deconvolution. *Astronomy and Astrophysics*, 528:A31+, 2011.
- [61] A. Leonard, F.-X. Dupé, and J.-L. Starck. A compressed sensing approach to 3D weak lensing. *Astronomy and Astrophysics*, 539:A85, 2012.
- [62] A. Leonard, F. Lanusse, F.-X. Dupé, and J.-L. Starck. *in prep.*, 2012.