Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **401,** 705–726 (2010) doi:10.1111/j.1365-2966.2009.15609.x

ASKI: full-sky lensing map-making algorithms

C. Pichon,[1,2,3★] E. Thiébaut,[1,2] S. Prunet,[1] K. Benabed,[1] S. Colombi,[1] T. Sousbie[1,4] and R. Teyssier[1,3,5]

[1]*Institut d'astrophysique de Paris & UPMC (UMR 7095), 98 bis boulevard Arago, 75 014, Paris, France*
[2]*Observatoire de Lyon (UMR 5574), 9 avenue Charles André, F-69561 Saint Genis Laval, France*
[3]*CEA/IRFU/SAP, l'Orme des Merisiers, 91170, Gif sur Yvette, France*
[4]*Tokyo University, Physics Dept 7-3-1 Hongo Bunkyo-ku, JP Tokyo 113-0033, Japan*
[5]*Institute für Theoretische Physik, Universität Zürich, Winterthurerstrasse 190 CH-8057 Zürich, Switzerland*

Accepted 2009 August 25. Received 2009 August 21; in original form 2008 December 26

ABSTRACT

Within the context of upcoming full-sky lensing surveys, the edge-preserving non-linear algorithm ASKI (*All-Sky κ Inversion*) is presented. Using the framework of Maximum A Posteriori inversion, it aims at recovering the *optimal* full-sky convergence map from noisy surveys with masks. ASKI contributes two steps: (i) CCD images of possibly crowded galactic fields are deblurred using *automated* edge-preserving deconvolution; (ii) once the reduced shear is estimated using standard techniques, the partially masked convergence map is also inverted via an edge-preserving method.

The efficiency of the deblurring of the image is quantified by the relative gain in the quality factor of the reduced shear, as estimated by SExtractor. Cross-validation as a function of the number of stars removed yields an automatic estimate of the optimal level of regularization for the deconvolution of the galaxies. It is found that when the observed field is crowded, this gain can be quite significant for realistic ground-based 8-m class surveys. The most significant improvement occurs when both positivity and edge-preserving $\ell_1 - \ell_2$ penalties are imposed during the iterative deconvolution.

The quality of the convergence inversion is investigated on noisy maps derived from the HORIZON-4π N-body simulation with a signal-to-noise ratio (S/N) within the range $\ell_{\rm cut} = 500$–2500, with and without Galactic cuts, and quantified using one-point statistics (S_3 and S_4), power spectra, cluster counts, peak patches and the skeleton. It is found that (i) the reconstruction is able to interpolate and extrapolate within the Galactic cuts/non-uniform noise; (ii) its sharpness-preserving penalization avoids strong biasing near the clusters of the map; (iii) it reconstructs well the shape of the PDF as traced by its skewness and kurtosis; (iv) the geometry and topology of the reconstructed map are close to the initial map as traced by the peak patch distribution and the skeleton's differential length; (v) the two-point statistics of the recovered map are consistent with the corresponding smoothed version of the initial map; (vi) the distribution of point sources is also consistent with the corresponding smoothing, with a significant improvement when $\ell_1 - \ell_2$ prior is applied. The contamination of B modes when realistic Galactic cuts are present is also investigated. Leakage mainly occurs on large scales. The non-linearities implemented in the model are significant on small scales near the peaks in the field.

Key words: gravitational lensing – methods: N-body simulations – methods: statistical – techniques: image processing – surveys – large-scale structure of Universe.

1 INTRODUCTION

In recent years, weak shear measurements have become a major source of cosmological information. By measuring the bending of the rays of light emerging from distant galaxies, one can gain some knowledge about the distribution of matter between the

★E-mail: pichon@iap.fr

emitter and us, and thus probe the properties and evolution history of dark matter (Bartelmann & Schneider 2001). This technique has led to significant results in a broad spectrum of topics, from measurements of the projected dark matter power spectrum (for the latest results, see Fu et al. 2008), 3D estimation of the dark matter spectrum (Kitching et al. 2006), studies of the higher order moments of the dark matter distribution, selection of source candidates for subsequent follow-ups (Schirmer et al. 2007) and the reconstruction of the mass distribution from small (Jee et al. 2007) to large scales (Massey et al. 2007). In view of these successes, numerous surveys have been planned specifically to use this probe either from ground-based facilities (e.g. VLT Survey Telescope-the Kilo-Degree Survey (VST-KIDS),[1] Dark Energy Survey (DES),[2] the Panoramic Survey Telescope and Rapid Response System (Pann-STARRS),[3] Large Synoptic Survey Telescope (LSST)[4]) or from space-based observatories (EUCLID,[5] Supernova Acceleration Probe (SNAP)[6] and The Joint Dark Energy Mission (JDEM)[7]). More generally, it is clear that weak lensing will be a major player in the future, as it has been identified by different European and US working groups as one of the most efficient way of studying the properties of dark energy.[8] Data processing is an important issue in the exploitation of weak lensing of distant galaxies. The signal comes from the excess alignment of the ellipticities of the observed galaxies. Assuming one can ignore or deal with spurious alignments due to intrinsic effects (Pichon & Bernardeau 1999; Aubert, Pichon & Colombi 2004; Hirata & Seljak 2004), or due to spurious lensing effects (Bridle & Abdalla 2007), the weak lensing signal will thus come from a small statistically coherent ellipticity on top of the random one of each object. Any result obtained with weak lensing on distant galaxies is thus conditioned by the quality with which shape parameters of the galaxies are recovered. This issue has of course been raised by the weak lensing community and tackled by the SHear Testing Program working group (Heymans 2006; Massey 2007) whose efforts have allowed for a fair comparison of the existing techniques. Schematically, the measurement of the shape parameters of the galaxies can be seen as a two-step process. First, one must correct for the non-idealities of the images due to atmospherical seeing (for ground-based telescopes), and telescope and camera aberrations. Indeed, these effects translate into an asymmetrical beam, which is varying between two images, and even possibly in the field of one image. Typically, the asymmetry induced by the instrumental response is much larger than the ellipticity to be measured. After this preprocessing step, a shape determination algorithm can be applied, and some estimation of the ellipticity of the object recovered. Stars, defects in the images and objects too close to each other after deconvolution have to be removed from the final catalogue so as to avoid contamination from erroneous shape measurements.

After these operations, one obtains a catalogue of position and shape parameters. Many techniques exist for recovering the weak shear signal from this catalogue. For example, a lot of efforts have been devoted to the measurement of the shear two-point functions. The most used method is the two-point functions; however, the measurement of the so-called *Mass Aperture* averaged two-point function, which is the result of the convolution of the shear two-point functions by a compensated filter (Schneider, van Waerbeke, Kilbinger & Mellier 2002), is becoming the preferred method (Fu et al. 2008). This scheme includes the separation between the curl-free convergence-field two-point function, and the residual curl mode that can arise from incomplete point spread function (PSF) correction or intrinsic galaxy alignment (Crittenden et al. 2002). For three-point functions, different resummation schemes have been proposed, either using direct measurement of the shear (Bernardeau, Mellier & van Waerbeke 2002; Benabed & Scoccimarro 2005) or using the *Mass Aperture* filter (Takada & Jain 2003; Kilbinger & Schneider 2005).

Other applications (source detection and fit, some tomography algorithms) call for an estimation of the map of the convergence field. A convergence map can also be used to measure the two- and three-point functions as well, even if, as we will see later, this is not optimal. For these reasons a lot of work has already been devoted to the reconstruction of the convergence map (Bartelmann et al. 1996; Seitz, Schneider & Bartelmann 1998; van Waerbeke, Bernardeau & Mellier 1999). The problem in this reconstruction lies in the inversion of the non-local equations linking the convergence field κ, and the ellipticities of the galaxies, while controlling the noise and avoiding pollution from the spurious curl modes. Moreover, even assuming that the ellipticity catalogue was a noise-free estimation of a curl-free underlying shear, the inversion could only be exact up to a global translation given the functional form of the equation. Thus, Bayesian techniques that use a priori properties on the solution to regularize the inversion problem are well suited to the reconstruction of κ. Previous works on the topic have explored different sets of a priori and regularization techniques (Marshall et al. 2002; Starck, Pires & Refregier 2006; Bridle et al. 1998; Seitz, Schneider & Bartelmann 1998). The primary goal of those investigations being the measurement of the mass distribution in clusters, most of them are dealing only with finite regions of the sky. For the same reason, those papers have been extended to include strong-lensing effects that can be observed around the cluster whose mass is being reconstructed using their lensing effect (Bradac et al. 2005; Cacciato et al. 2006; Halkola, Seitz & Pannella 2006; Jee et al. 2007).

In this paper, we will focus on the optimal reconstruction of the κ *field* from very large, and possibly full-sky, maps of the sky. We will thus only be interested in the weak lensing regime including the onset of the quasi-linear regime, where the non-linearities of the relation linking the ellipticities of the galaxies to the shear cannot be safely neglected. We will propose a self-calibrated regularization technique that can be compared to multi-resolution methods or wavelet approach (Starck et al. 2006; Abrial et al. 2008), and use a $\ell_1 - \ell_2$ regularization scheme to perform a sharp feature preserving inversion. One of the biggest issues we will have to cope with is the incomplete coverage of the sky. We will show how our technique can deal with irregular coverage and masked portions of the sky.

Specifically, Section 2 shows how self-calibrated non-parametric $\ell_1 - \ell_2$ deblurring can *improve* the construction of reduced shear, hence convergence maps. Section 3 describes the model for the reduced shear, the corresponding inverse problem and the optimization procedure. Section 4 investigates the quality of the global reconstruction; in particular, it probes the asymmetry/kurtosis of the recovered maps, its topology (total length and differential length of the skeleton), the recovered power spectra, the point source catalogue with and without galactic star cut. The leaking of B modes

[1] http://www.astro-wise.org/projects/KIDS/
[2] https://www.darkenergysurvey.org/
[3] http://pan-starrs.ifa.hawaii.edu/
[4] http://www.lsst.org/
[5] http://www.dune-mission.net/
[6] http://snap.lbl.gov/
[7] http://universe.nasa.gov/program/probes/jdem.html
[8] See on the European side http://www.stecf.org/coordination/ and on the US side http://www.nsf.gov/mps/ast/aaac.jsp and http://www.nsf.gov/mps/ast/detf.asp

induced by the Galactic cut is also investigated. Finally, Section 5 discusses implications for upcoming full-sky surveys and wraps up.

Appendix A describes the star removal algorithm (implemented for the cross-validation estimation of the optimal level of smoothing required to deconvolve the crowded images); Appendix B details the κ inverse problem on the sphere while Appendix C derives the local plane corresponding approximation. Appendix D describes the construction of realistic κ maps from large N-body simulations.

2 DEBLURRING OF CROWDED FIELDS

The first step involved in reconstructing a full-sky map of the convergence on the sky requires estimating ellipticity and orientation maps from wide angle CCD images of large patches of the sky. Whether the experiment is ground-based, or space-born, it is advisable to correct for the effect of the instrumental response, in particular when mapping more crowded regions closer to the Galactic plane. Indeed, the PSF-induced partial overlapping of galaxies within the field of view will bias the estimation of the reduced shear. What we will describe here would correspond to a method belonging to the 'orange' quadrant of the classification proposed in table 3 of Massey (2007). Current methods have been designed for deblurring of isolated objects and are consequently less efficient in deblurring blended objects. As a first step towards building a full-sky map maker, let us therefore address the issue of deblurring crowded fields via regularized non-parametric model fitting, and assess its efficiency in the weak lensing context.

In particular, we will show that cross-validation as a function of the number of stars removed yields an automatic estimate of the optimal level of regularization for the deconvolution of the galaxies. When the observed field is crowded, this gain can be quite significant for realistic ground-based 8-m class surveys. The most significant improvement occurs when both positivity and edge-preserving $\ell_1 - \ell_2$ penalties are imposed during the iterative deconvolution.

2.1 Deblurring as an inverse problem

2.1.1 Regularized solution

Since observed objects are incoherent sources, the observed image depends linearly on the sky brightness distribution:

$$y(\boldsymbol{\omega}) = \int h(\boldsymbol{\omega}, \boldsymbol{\omega}')\, x(\boldsymbol{\omega}')\, \mathrm{d}\boldsymbol{\omega}' + e(\boldsymbol{\omega})\,,$$

where $y(\boldsymbol{\omega})$ is the observed distribution in the direction $\boldsymbol{\omega}$, $h(\boldsymbol{\omega}, \boldsymbol{\omega}')$ is the atmospheric and instrumental PSF which is the distribution of observed light in the direction ω due to light coming from direction $\boldsymbol{\omega}'$, $x(\boldsymbol{\omega}')$ is the true sky brightness distribution and $e(\boldsymbol{\omega})$ is the noise. After discretization

$$\boldsymbol{y} = \mathbf{H} \cdot \boldsymbol{x} + \boldsymbol{e}\,, \tag{1}$$

where $\boldsymbol{y}$ is the vector of pixel intensities in the observed image (the data), $\mathbf{H}$ is the matrix which accounts for the atmospheric and instrumental blurring, $\boldsymbol{x}$ is the (discretized or projected on to a basis of functions) object brightness distribution and $\boldsymbol{e}$ accounts for the errors (pixelwise noise and modelization approximations). Deblurring requires estimating the best sky brightness distribution given the data. Since the atmospheric and instrumental PSF results in a smoother distribution than the true one, it is well known that de-blurring is an ill-conditioned problem (Richardson 1972; Skilling, Strong & Bennett 1979; Tarantola & Valette 1982; Pichon & Thiébaut 1998; Pichon et al. 2001). In other words, straightforward deblurring by applying $\mathbf{H}^{-1}$ to the data $\boldsymbol{y}$ would result in uncontrolled amplification of noise: a small change in the input data would yield unacceptably large artefacts in the solution. Regularization must be used to overcome ill-conditioning of this inverse problem. This is achieved by using additional prior constraints such as requiring that the solution be as smooth as possible, while being still in statistical agreement with the data and while imposing that the brightness distribution is positive. Following this prescription, the Maximum A Posteriori (MAP) solution $\boldsymbol{x}_\mu$ is the one which minimizes an objective function $\mathcal{Q}(\boldsymbol{x})$:

$$\boldsymbol{x}_\mu = \arg\min_{\boldsymbol{x}\geq 0} \mathcal{Q}(\boldsymbol{x})\,, \quad \text{with: } \mathcal{Q}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}) + \mu\, \mathcal{R}(\boldsymbol{x})\,, \tag{2}$$

where $\mathcal{L}(\boldsymbol{x})$ is a likelihood penalty which enforces agreement of the model with the data, $\mathcal{R}(\boldsymbol{x})$ is a *regularization* penalty which enforces prior constraints set on the model and $\mu > 0$ is a so-called *hyper-parameter* which allow the tuning of the relative weight of the prior with respect to the data. Hence, the MAP solution is a compromise between what can be inferred from the data alone and prior knowledge about the parameters of interest. Assuming Gaussian statistics for the errors $\boldsymbol{e}$ in equation (1), the likelihood penalty writes

$$\mathcal{L}(\boldsymbol{x}) = (\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y})^{\mathrm{T}} \cdot \mathbf{W} \cdot (\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y})\,, \tag{3}$$

where the weighting matrix $\mathbf{W}$ is equal to the inverse of the covariance matrix of the errors: $\mathbf{W} \equiv \mathrm{Cov}(\boldsymbol{e})^{-1}$.

The most effective regularization for ill-conditioned problems such as deconvolution of blurred images consists in imposing a smoothness constraint (Thiébaut 2005). Then, the regularization penalty writes

$$\mathcal{R}(\boldsymbol{x}) = \sum_j \phi(\Delta x_j)\,, \tag{4}$$

where Δx_j is the local gradient of $\boldsymbol{x}$ and ϕ is some cost function. The local gradient of $\boldsymbol{x}$ can be approximated by finite differences: $\Delta\boldsymbol{x} = \mathbf{D} \cdot \boldsymbol{x}$ where $\mathbf{D}$ is a linear finite difference operator. For instance, in 1D: $\Delta x_j = (\mathbf{D} \cdot \boldsymbol{x})_j = x_{j+1} - x_j$. To enforce smoothness, the cost function ϕ must be an increasing function of the magnitude of its argument. Very common choices for ϕ are: the ℓ_2 norm, the ℓ_1 norm or an $\ell_1 - \ell_2$ norm. For our deblurring problem, we have considered different priors (quadratic or $\ell_1 - \ell_2$ smoothness) possibly with an additional positivity constraint. We have used generalized cross-validation (GCV; Wahba 1990) applied to the circulant approximation of the quadratic problem to estimate the optimal regularization level μ. These different possibilities and their effects on the recovered images are discussed in details in what follows.

Finally, to solve for the constrained optimization problem (2), we used the VMLMB algorithm from OPTIMPACK (Thiébaut 2002). VMLMB (for *Variable Metric, Limited Memory, Bounded*) makes use of a BFGS (Nocedal & Wright 2006) update of the approximation of the Hessian (matrix of second partial derivatives) of $\mathcal{Q}(\boldsymbol{x})$ to derive a step to improve the parameters at every iteration. This strategy only requires computing the objective function, $\mathcal{Q}(\boldsymbol{x})$, and its gradient (partial derivatives) $\nabla_{\boldsymbol{x}}\mathcal{Q}(\boldsymbol{x})$ with respect to the parameters $\boldsymbol{x}$. The BFGS update is limited to a few last steps so that the memory requirements remain modest, that is a few times the number of sought parameters, and the algorithm can be applied to solve very large problems (in our case, there are as many parameters as the number of pixels in the sought image). Finally, VMLMB accounts for bound constraints by means of gradient projections (Nocedal & Wright 2006). For a convex penalty $\mathcal{Q}(\boldsymbol{x})$, VMLMB is guaranteed to converge to the unique feasible minimum of $\mathcal{Q}(\boldsymbol{x})$ which satisfies the bound constraints; for a non-convex penalty, VMLMB being based

on a descent strategy, it will find a local minimum depending on the initial set of parameters.

2.1.2 Quadratic regularization and Wiener proxy

Using the finite difference operator $\mathbf{D}$ and an ℓ_2 norm for the regularization and ignoring for the moment the positivity constraint, the MAP solution is the minimum of a quadratic penalty which simply involves solving a (huge) linear problem:

$$\begin{aligned}\boldsymbol{x}_\mu &= \arg\min_{\boldsymbol{x}} \Big\{ (\mathbf{H}\cdot\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}\cdot\mathbf{W}\cdot(\mathbf{H}\cdot\boldsymbol{x}-\boldsymbol{y}) \\ &\qquad +\mu\,(\mathbf{D}\cdot\boldsymbol{x})^{\mathrm{T}}\cdot(\mathbf{D}\cdot\boldsymbol{x})\Big\} \\ &= (\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\cdot\mathbf{H}+\mu\,\mathbf{D}^{\mathrm{T}}\cdot\mathbf{D})^{-1}\cdot\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\cdot\boldsymbol{y}\,, \end{aligned} \tag{5}$$

providing the Hessian matrix $\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\cdot\mathbf{H}+\mu\,\mathbf{D}^{\mathrm{T}}\cdot\mathbf{D}$ is non-singular, which is generally the case for $\mu>0$. Owing to the large size of the matrices involved in this equation (there are as many unknown as the number of pixels), the linear problem has to be iteratively solved (by a limited memory algorithm such as VMLM) unless it can be diagonalized as explained below. The solution, equation (5), involves at least one parameter, μ, which needs to be set to the correct level of regularization: too low would give a solution plagued by lots of artefacts due to noise amplification, and too high would result in an oversmoothed solution with small details blurred. The optimal level of smoothing can be computed by GCV by minimizing with respect to μ the function (Golub, Heath & Wahba 1979; Wahba 1990)

$$\mathrm{GCV}(\mu) = \frac{\left(\mathbf{A}_\mu\cdot\boldsymbol{y}-\boldsymbol{y}\right)^{\mathrm{T}}\cdot\mathbf{W}\cdot\left(\mathbf{A}_\mu\cdot\boldsymbol{y}-\boldsymbol{y}\right)}{\left[1-\mathrm{tr}(\mathbf{A}_\mu)/N\right]^2}\,, \tag{6}$$

where N is the number of data (size of $\boldsymbol{y}$) and $\mathbf{A}_\mu=\nabla_{\boldsymbol{y}}(\mathbf{H}\cdot\boldsymbol{x}_\mu)$ is the so-called *influence matrix*, in our case

$$\mathbf{A}_\mu = \mathbf{H}\cdot\left(\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\cdot\mathbf{H}+\mu\,\mathbf{D}^{\mathrm{T}}\cdot\mathbf{D}\right)^{-1}\cdot\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\,. \tag{7}$$

Computing the value of GCV (μ) involves (i) solving the problem to find the regularized solution $\boldsymbol{x}_\mu$ and computation of $\mathbf{A}_\mu\cdot\boldsymbol{y}=\mathbf{H}\cdot\boldsymbol{x}_\mu$; (ii) estimation of the trace of $\mathbf{A}_\mu$ perhaps by using Monte Carlo methods (Girard 1989) since the influence matrix is very large. The computational cost of Stages (i) and (ii) is similar to that of a few solvings of the quadratic problem. Since this has to be repeated for every different value of the regularization level, finding the optimal value of μ by means of GCV can be very time consuming unless the problem can be approximated by a diagonal quadratic problem (for which matrix inversions are both fast and trivial).

For this purpose, we introduce the proxy problem corresponding to white noise and circulant approximations of the operators $\mathbf{H}$ (convolution by the PSF) and $\mathbf{D}$ (finite differences). Then the weighting matrix becomes

$$W_{i,j}=\delta_{i,j}/\sigma^2\,,\quad \text{where } \sigma^2=\mathrm{Var}(n_i),$$

where $\sigma^2=\mathrm{Var}\,(e_i)$ is the variance of the noise. In the special case where the PSF is *shift-invariant*, $\mathbf{H}$ is a convolution operator which can be approximated by a block Toeplitz with Toeplitz block matrix that can be computed very quickly by means of Fast Fourier Transforms (FFTs):

$$\mathbf{H}\cdot\boldsymbol{x}\simeq\mathbf{F}^{-1}\cdot\mathrm{diag}(\mathbf{F}\cdot\boldsymbol{h})\cdot(\mathbf{F}\cdot\boldsymbol{x})\,, \tag{8}$$

where $\boldsymbol{h}$ is the PSF (the first row of $\mathbf{H}$), $\mathbf{F}$ is the forward DFT operator and diag ($\boldsymbol{v}$) is the diagonal matrix with its diagonal given by the vector $\boldsymbol{v}$. This discrete convolution equation assumes that $F_{u,j}=\exp(-2\,\mathrm{i}\,\pi\,\sum_n u_n\,j_n/N_n)$ where N_n is the length of the n th dimension, $j_n=0,\ldots,N_n-1$ and $u_n=0,\ldots,N_n-1$ are the indices of the position and discrete Fourier frequency along this dimension. In this case, the inverse DFT is simply $\mathbf{F}^{-1}=\mathbf{F}^{\mathrm{H}}/N_{\mathrm{tot}}$ with N_{tot} the total number of elements in $\boldsymbol{x}$ and the H exponent standing for the conjugate transpose. With these approximations and definitions of the DFT, the likelihood term writes

$$\mathcal{L}(\boldsymbol{x})=\frac{1}{\sigma^2}\,\|\mathbf{H}\cdot\boldsymbol{x}-\boldsymbol{y}\|^2\simeq\frac{1}{N_{\mathrm{tot}}\,\sigma^2}\,\sum_u|\hat{h}_u\,\hat{x}_u-\hat{y}_u|^2\,, \tag{9}$$

where $\hat{h}_u$ is the transfer function (the DFT of the PSF) and $\hat{y}_u$ and $\hat{x}_u$, respectively, the DFT of the data and of the sought image. Note that the exact normalization factor, here $1/N_{\mathrm{tot}}$, depends on the particular definition of the DFT.

Similarly, ignoring edges effects, the finite difference operator $\mathbf{D}$ along n th direction can be approximated by

$$\mathbf{D}_n\cdot\boldsymbol{x}\simeq\mathbf{F}^{-1}\cdot\mathrm{diag}(\hat{\boldsymbol{d}}_n)\cdot(\mathbf{F}\cdot\boldsymbol{x})\,, \tag{10}$$

where $\hat{\boldsymbol{d}}_n$ is the DFT of the first row of $\mathbf{D}_n$; then the quadratic regularization writes

$$\mathcal{R}(\boldsymbol{x})=\|\mathbf{D}\cdot\boldsymbol{x}\|^2=\sum_n\|\mathbf{D}_n\cdot\boldsymbol{x}\|^2\simeq\frac{1}{N_{\mathrm{tot}}}\,\sum_u r_u\;|\hat{x}_u|^2\,, \tag{11}$$

with

$$r_u=\sum_n|\hat{d}_{n,u}|^2=4\,\sum_n\sin^2\left(\frac{\pi\,u_n}{N_n}\right)\,, \tag{12}$$

for the first-order finite differences and our choice for the DFT. Note that any $r_u\geq0$ being an increasing function of the length $|u|$ of the spatial frequency could be used instead and would result in imposing a smoothness constraint although with a different behaviour. Finally, putting all these circulant approximations together, the quadratic problem to solve is diagonalized in the DFT space and trivially solved to give the DFT of the MAP solution:

$$\hat{x}_{\mu,u}=\frac{\hat{h}_u^\star\,\hat{y}_u}{|\hat{h}_u|^2+\mu\,\sigma^2\,r_u}\,, \tag{13}$$

the asterisk exponent denoting the complex conjugate. Note that this circulant approximation of the solution is very fast to compute as it involves just a few FFTs. This expression of the MAP solution is very similar to what would give the Wiener filter which would be exactly achieved by setting the term $\mu\,r_u$ equal to the reciprocal of the expected image power spectrum in equation (13). Since, in our case, the image power spectrum is unknown a priori, we have to choose the *regularization shape* r_u and derive the optimal level of smoothing, for instance, by means of GCV. Thanks to the circulant approximation made here, GCV criterion is now very easy to compute as

$$\mathbf{A}_\mu\simeq\mathbf{F}^{-1}\cdot\mathrm{diag}(\hat{\boldsymbol{a}}_\mu)\cdot\mathbf{F}\,,\quad\text{with : }\hat{a}_{\mu,u}=\frac{|\hat{h}_u|^2}{|\hat{h}_u|^2+\mu\,\sigma^2\,r_u}\,,$$

and $\mathrm{tr}(\mathbf{A}_\mu)=\sum_u\hat{a}_{\mu,u}/N_{\mathrm{tot}}$, hence

$$\mathrm{GCV}(\mu)=\frac{N_{\mathrm{tot}}\,\sum_u t_{\mu,u}^2\,|\hat{y}_u|^2}{\sigma^2\,[\sum_u t_{\mu,u}]^2}\,, \tag{14}$$

with

$$t_{\mu,u}=1-\hat{a}_{\mu,u}=\frac{\mu\,\sigma^2\,r_u}{|\hat{h}_u|^2+\mu\,\sigma^2\,r_u}\,. \tag{15}$$

In practice, for the optimization of equation (2), equation (13) is taken as a starting point together with the choice of μ given by the minimum of equation (14). Then, the optimization of equation (2) is carried with possibly non-stationary weights, while iterating back and forth between model and data space.

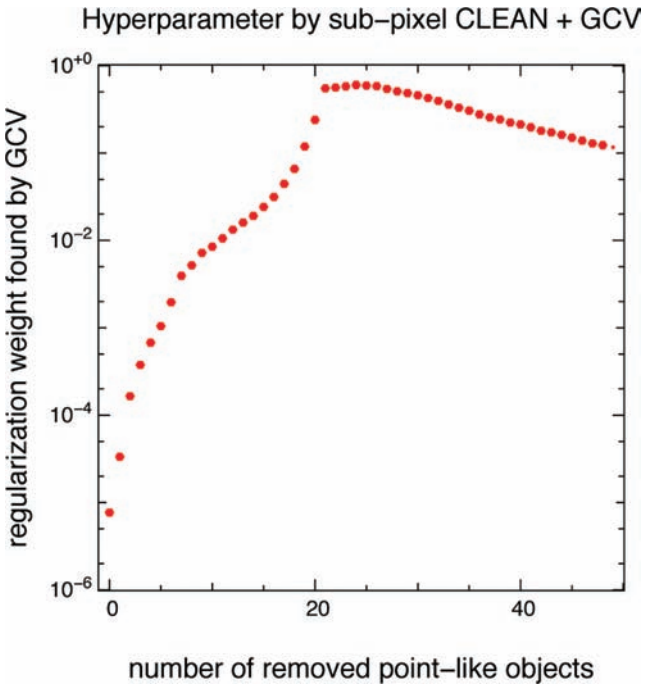

Figure 1. Hyper-parameter chosen by GCV as a function of the number of stars removed by our star removal algorithm (Appendix A). Note that this curve reaches a maximum corresponding to the moment when all the stars have been removed. Indeed, stars correspond to high frequency-correlated signal, while the wings of galaxies (for which the core has been erroneously removed) also give rise to such signal. In between, when all stars have been removed, while no galaxy has yet been deprived of its core, the amount of correlated high-frequency signal reaches a minimum, or equivalently the GCV estimated value of μ reaches a maximum.

2.1.3 *Crowded fields and star removal*

Even though the estimation of the ellipticities does not require *per se* the deconvolution of the galaxies, it is shown below that this estimation is significantly improved by deconvolution when the fields of view are crowded and polluted by foreground stars: indeed galaxies and stars overlap less when deconvolved, which reduces the fraction of erroneous measurements. Unfortunately, when these stars are present, they significantly bias the estimation of the hyper-parameter, μ, since stars correspond to high frequency-correlated signal which leads to an underestimation of the optimal level of smoothing (for the galaxies) by cross-validation. This is best seen in Fig. 1 which displays the evolution of the hyper-parameter which minimizes GCV as a function of the number of stars removed by our star removal algorithm (see Appendix A). Interestingly, it suggests that GCV could be used as a classifier.

2.1.4 $\ell_1 - \ell_2$ *penalty and positivity*

The drawback of using a quadratic (ℓ_2) norm in the regularization is that it tends to oversmooth the regularized map especially around sharp features as point-like sources (i.e. stars) and the core of galaxies. This is because the regularization prevents large intensity differences between neighbouring pixels and result in damped oscillations (the Gibbs effect). Such ripples hide any faint details in the vicinity of sharp structures. To avoid this, it would be better to use a regularization which smoothes out small local fluctuations of the sought distribution (here the deblurred image), presumably due to noise, but let larger local fluctuations arise occasionally (see Aubert & Kornprobst 2008 and references therein). This can be achieved by using a $\ell_1 - \ell_2$ cost function ϕ in equation (4). A possible $\ell_1 - \ell_2$ sparse cost function is (Mugnier, Fusco & Conan 2004)

$$\phi(r) \equiv 2\,\varepsilon^2 \left[\left|\frac{r}{\varepsilon}\right| - \log\left(1 + \left|\frac{r}{\varepsilon}\right|\right)\right] . \tag{16}$$

For a small, respectively large, pixel differences r, $\phi(r)$ has the following behaviour:

$$\phi(r) \sim \begin{cases} r^2 & \text{when } |r| \ll \varepsilon\,, \\ 2\,|\varepsilon\, r| & \text{when } |r| \gg \varepsilon\,, \end{cases}$$

which shows that, as required, the $\ell_1 - \ell_2$ penalty behaves quadratically for small *residuals*rs (in magnitude and w.r.t. ε) and only linearly for large rs. The derivative, needed for the optimization algorithm, of the $\ell_1 - \ell_2$ penalty writes

$$\phi'(r) = \frac{2\,\varepsilon\, r}{\varepsilon + |r|}\,.$$

An additional possibility to improve the restitution of faint details with level close to that of the background is to apply a strict positivity constraint. This is achieved by using VMLMB, a modified limited memory variable metric method (Thiébaut 2002), which imposes simple bound constraints by means of gradient projection. This yields a reduction of aliasing by bounding the allowed region of parameter space which can be explored during the optimization.

2.2 Numerical experiments

The public package SkyMaker (Erben et al. 2001) was used to generate galactic and stellar fields from ellipticity and magnitude catalogues. Table 1 summarizes the main parameter corresponding to the VLT with a Visible Multi-Object Spectrograph (VIMOS) instrument, a worse case situation compared to upcoming space missions.

A regular grid of 12 × 12 galaxies of magnitude 20 with random orientation is produced twice (with the same random seed), one corresponding to a fixed seeing and a given exposure time, while the other assuming zero noise and zero seeing for a set of 512 × 512 pixels images (see Fig. 2).

Table 1. SkyMaker parameters used to generate the VIMOS/VLT images.

Object	Value
Gain (e-/ADU)	30.11
Full well capacity in e-	300 000
Saturation level (ADU)	60 000
Read-out noise (e-)	1.3
Magnitude zero-point (ADU per second)	21.254
Pixel size in arcsec	0.2
Number of microscanning steps	1
SB (mag arcsec^{-2}) at 1 arcmin from a 0-mag star	16.0
Diameter of the primary mirror (m)	8.0
Obstruction diameter from second mirror in metre	2.385
Number of spider arms (0 = none)	4
Thickness of the spider arms (in mm)	5.0
Pos. angle of the spider pattern	45.0
Average wavelength analyzed (μm)	0.80
Back. surface brightness (mag arcsec^{-2})	21.5
Nb of stars/□ brighter than MAG_LIMITS	1e5
Slope of differential star counts (dexp mag^{-1})	0.3
Stellar magnitude range allowed	12.0,19.0

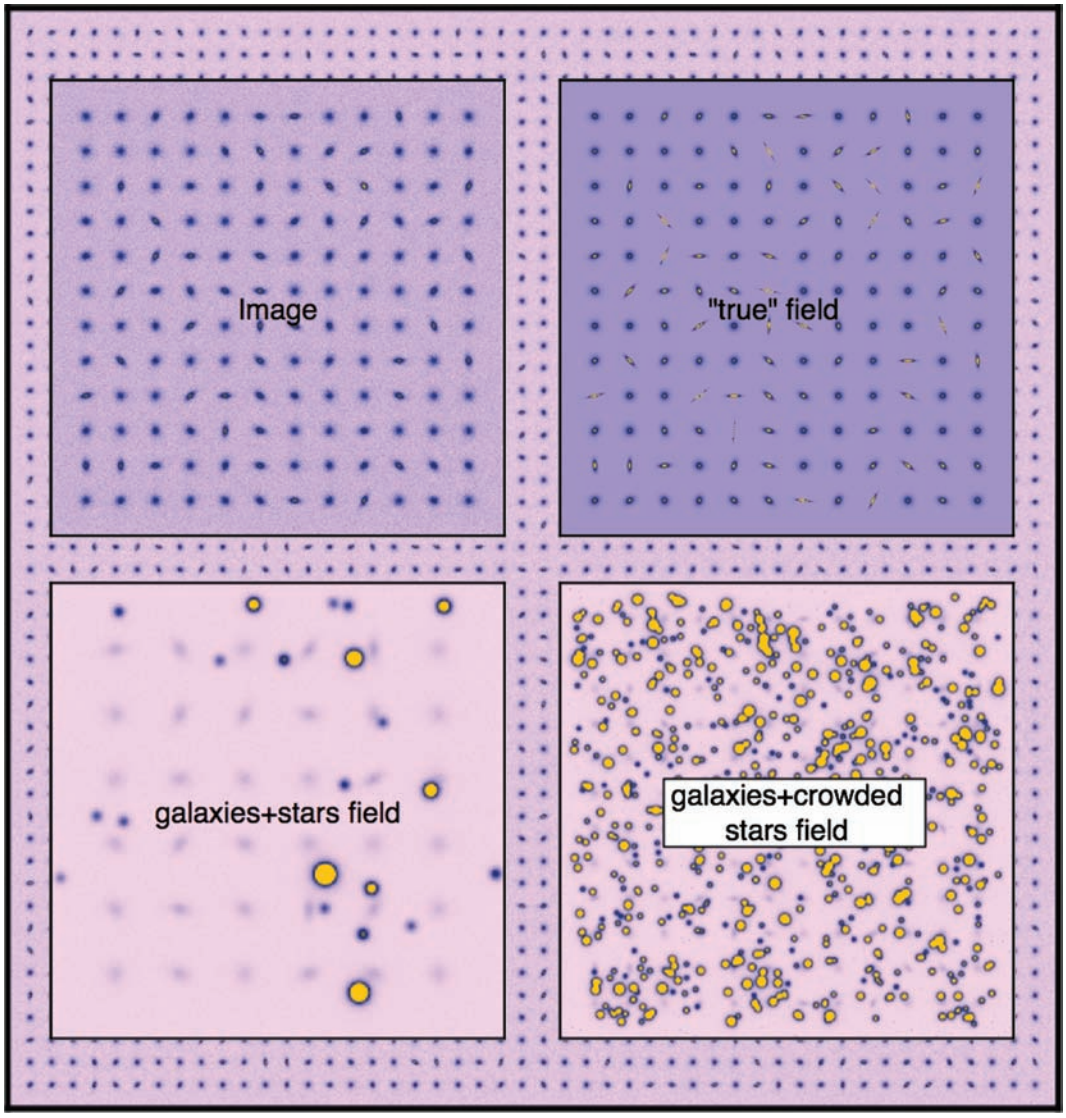

Figure 2. An example of virtual fields generated with SKYMAKER to be fed to SEXTRACTOR before and after deconvolution using the different regularizations described in the text. From top to bottom and left to right: a galaxy field image and the corresponding 'true' field, galaxies field with stars and a crowded galaxy field with stars (10^6 stars arcmin^{-2}). The exposure time is 10 s and the seeing is 1 arcsec for the VLT with VIMOS. The background field corresponds to the actual size of the corresponding observed images.

The background level and the amplitude of the background noise are first estimated automatically from the histogram of the pixel values and fed to SEXTRACTOR (Bertin & Arnouts 1996) which then estimates the position, the flux, the orientation and the ellipticity for all the galaxies in the field. Here, the ellipticity is defined as $1 - b/a$, where a and b are the long and short axes. This procedure is reproduced 50 times with different realizations. The measured and the recovered ellipticity are compared, together with flux and orientation for all the galaxies in the field. In this set of simulations, the prior knowledge of the position of the galaxy is used to minimize errors which might arise while using SEXTRACTOR: the recovered galaxy is chosen to be that which is closest to the known input position. The median and interquartile of the error (difference between the 'true' and recovered) in ellipticity versus the ellipticity is computed for a range of exposure time; this procedure is iterated for the three deconvolution techniques used in this paper (Wiener, ℓ_2 with positivity, $\ell_1 - \ell_2$ with positivity). An example of such a plot is shown in Fig. 3. Clearly, the bias in the recovered ellipticity increases with the ellipticity and the amount of noise in the image (via poorer seeing or shorter exposure time). As expected, the Wiener deconvolution is the least efficient of the three methods, since the linear penalty does not avoid some level of Gibbs ringing. In contrast, the ℓ_2 penalty with positivity avoids partially such ringing, while the $\ell_1 - \ell_2$ penalty works best at recovering the input eccentricity with a consistent level of bias below 10 per cent for an ellipticity in the range [0.1, 0.8[. Note that this bias is relative, not absolute. If an alternative shear estimator that does not consider deconvolution is accurate to a level of, say, 1 per cent, the expected bias after deconvolution will be below 0.1 per cent.

Interestingly, there is also a residual bias (even for longer exposure times) for small ellipticity galaxies, which arises because noise-induced departure from sphericity is amplified by the deconvolution. Note that the Wiener deconvolution is significantly faster than the iterative deconvolution with positivity (with ℓ_2 or $\ell_1 - \ell_2$ penalties). Positivity improves significantly the deconvolution, but will depend critically on the ability to estimate the background. In

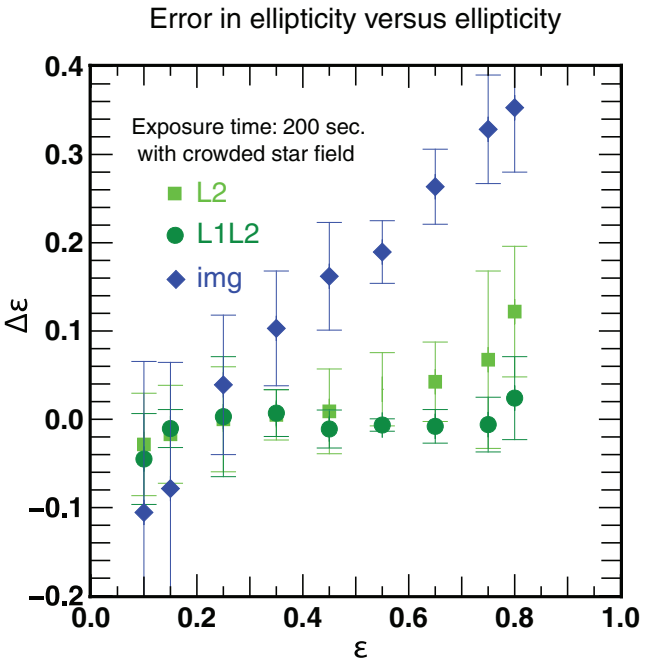

Figure 3. The error in ellipticity as a function of the ellipticity (measured by SExtractor) for a set of 50 images (such as those shown in Fig. 2) either directly on the image (medium diamonds), deconvolved with ℓ_2 gradient penalty function with enforced positivity (light squares) and $\ell_1 - \ell_2$ gradient penalty function with positivity (dark circles). For each set, the ellipticity is also measured directly on the raw image. Note that, as expected, the error on the bias is largest for circular galaxies, since deconvolution will tend to overamplify departure from circular symmetry.

the present simulations, the level of background is automatically estimated while looking at the histogram of the pixels. Finally, the $\ell_1 - \ell_2$ regularization significantly improves the restoration of fields of stars and galaxies, because the stars and the cores of galaxies are very sharp. These non-linear iterative methods are slower than the Wiener filtering, but can account at no extra cost for non-uniform noise, or saturation and masking. Their convergence can be considerably boosted when they are initiated by the Wiener solution.

For any such plot, two numbers are defined which summarize the trend. The mean error (averaged over the various ellipticities), $\bar{\epsilon}$, and the mean of the interquartile, $\Delta\bar{\epsilon}$, were measured. The quality factor, QF, is defined to be the ratio of the sum of this mean error and the mean interquartile for the image without deblurring, divided by the sum of the mean error and the mean interquartile for the deconvolved image for the three techniques (Wiener, ℓ_2 and $\ell_1 - \ell_2$). This reads

$$QF_{\rm method} = \frac{\bar{\epsilon}_{\rm image} + \Delta\bar{\epsilon}_{\rm image}}{\bar{\epsilon}_{\rm method} + \Delta\bar{\epsilon}_{\rm method}}\,.$$

The evolution of the quality of the ellipticity measurement is traced versus seeing conditions and signal-to-noise ratio (exposure time) in two regimes: a galaxy-only field and a galactic field with a crowded star content where the number of stars per square degree reaches 10^5 stars arcmin^{-2}. These two regime represent high and low Galactic region, respectively. Fig. 4 displays the evolution of $QF_{\rm Wiener}$ (diamonds), QF_{ℓ_2} (triangles), and $QF_{\ell_1-\ell_2}$ (circles), as a function the exposure time of 1, 10, 100 and 1000 s, respectively, and two seeing conditions of 1.2 and 0.7 arcsec. No stars are present in the field on the left-hand panel of Fig. 4, whereas its right-hand panel displays the three QF estimators for a field with a realistic 10^5 stars per square degree. ASKI achieves efficient deblurring in this regime. It remains to be shown that regularized deconvolution obtained through (sparse) parametric local decomposition of both PSF and objects (as done e.g. with shapelet-based methods) can properly deblur blended objects.

Now that we have shown that the state-of-the-art automated positive edge-preserving deconvolution of deep sky images is mandatory to get good-quality shear estimates (most importantly in the context of crowded fields), let us conclude this section by a leap forward, and assume from now on that not only we have access to discrete

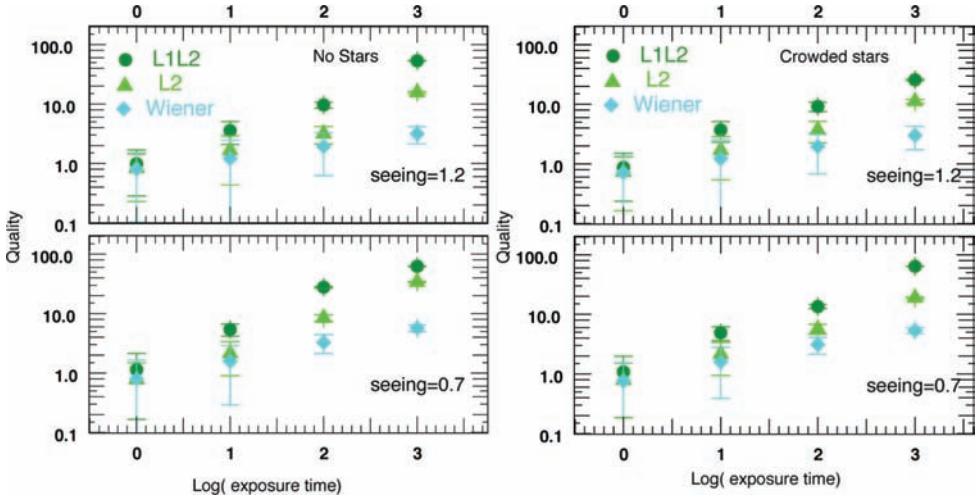

Figure 4. Left-hand panel: the relative error quality factor (see main text) as a function of the log exposure time for the three methods, respectively the Wiener filtering (diamonds), ℓ_2 gradient penalty function with enforced positivity (triangles) and $\ell_1 - \ell_2$ gradient penalty function with positivity (circles). Two seeing conditions are investigated, corresponding to a good (0.7 arcsec) and a fair (1.2 arcsec) seeing condition. These simulations assume that no star is present in the field, and correspond to a set of non-overlapping galactic discs with random orientation and magnitude 20 in V (see Fig. 2). The telescope setting corresponds to the VIMOS instrument on an 8-m VLT. Right-hand panel: the quality factor as a function of the log exposure time, but this time while allowing for stars in the field. The star count is 10^5 stars per arcmin^2. As discussed in the text, the penalty weight is estimated via GCV on a temporary image where all stars are automatically removed via blind cleaning as described in Appendix A. Here, the removal of stars is essential since the GCV hyper-parameter (which sets the level of smoothing in the deconvolved image) varies by orders of magnitudes in the process (see Fig. 1 for a discussion) and would be otherwise underestimated.

measurements of ellipticities over a significant fraction of the sky, but also this point-like process has been re-sampled. Indeed, since it is beyond the scope of this paper to carry out a full-sky deconvolution and reconstruction at the resolution of 0.7 arcsec (This would amount to about 10^{12} pixels!), it is assumed from now on that a full-sky catalogue of vector-reduced shear exists and that the interpolation/re-sampling of the corresponding map on a uniform grid over the sphere has been done, together with an estimate of the corresponding shot noise. In other words, we skip the critical step of optimal shear estimation, which has already been addressed by the STEP (Heymans 2006; Massey 2007) working group. In this paper, we extract the virtual catalogue from a state-of-the-art simulation (see below); we make use of the HEALPIX Pixelization (Górski et al. 1999), a hierarchical equi-surface and iso-latitude pixelization of the sphere, which was developed to analyze polarized cosmic microwave background (CMB)-type data.

3 A FULL-SKY MAP MAKER

3.1 The inverse problem

Our purpose is now to solve for the non-linear inverse problem of recovering the $\kappa(\hat{\boldsymbol{n}})$ map corresponding to a noisy incomplete measurement of the 2D field $(g_1(\hat{\boldsymbol{n}}), g_2(\hat{\boldsymbol{n}}))^{\mathrm{T}}$ of the ellipticity and orientation on the sphere (in the local tangent plane):

$$g_k(\hat{\boldsymbol{n}}) = \frac{\gamma_k(\hat{\boldsymbol{n}})}{1-\kappa(\hat{\boldsymbol{n}})} + e_k(\hat{\boldsymbol{n}}), \quad \text{for } k = 1 \text{ or } 2, \tag{17}$$

where $\hat{\boldsymbol{n}}$ is the sky direction, γ and κ are, respectively, the shear and the convergence, while $\boldsymbol{e}$ is a tensor field of the *errors* which accounts for the measurement noise (including the shot noise induced by the finite number of galaxies within that pixel) and model approximations.

3.1.1 Spherical formulation

On the sphere, the scalar field κ and the tensor field γ are linear functions of the *unknown* complex field $\boldsymbol{a} = \mathbf{Y}\cdot\boldsymbol{\kappa}$ whose coefficients are the spherical harmonic coefficients of κ. After discretization and using matrix notation, κ and γ write

$$\kappa \equiv \mathbf{K}\cdot\boldsymbol{a} \quad \text{and} \quad \gamma \equiv \mathbf{G}\cdot\boldsymbol{a}, \tag{18}$$

where $\mathbf{K} = \mathbf{Y}$ and $\mathbf{G} = {}_p\mathbf{Y}\cdot\mathbf{J}$, $\mathbf{Y}$ denoting the scalar spherical harmonics and ${}_p\mathbf{Y} = ({}_E\mathbf{Y}, {}_B\mathbf{Y})$ the parity eigenstates based on spin-2 spherical harmonics. These eigenstates are defined in such a way that

$$\gamma_1 \pm i\gamma_2 = -\sum_{\ell m}(a_{\ell,m,E} \pm i a_{\ell,m,B}) {}_{\pm2}\mathbf{Y}_{\ell m},$$

so that we have

$$\begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix} = \sum_{\ell,m}\begin{pmatrix}-\mathbf{W}^+_{\ell,m}\\ +i\,\mathbf{W}^-_{\ell,m}\end{pmatrix} a_{\ell,m,E} + \sum_{\ell,m}\begin{pmatrix}-i\,\mathbf{W}^-_{\ell,m}\\ -\mathbf{W}^+_{\ell m}\end{pmatrix} a_{\ell,m,B}$$

with $\mathbf{W}^{\pm}_{\ell,m} = ({}_2\mathbf{Y}_{\ell,m} \pm {}_{-2}\mathbf{Y}_{\ell,m})/2$. Here, $\mathbf{J}$ operates on $\boldsymbol{a}$ as

$$(\mathbf{J}\cdot\boldsymbol{a})_{\ell,m,E} = \sqrt{\frac{(\ell+2)(\ell-1)}{(\ell+1)\ell}}\, a_{\ell,m}, \tag{19}$$

$$(\mathbf{J}\cdot\boldsymbol{a})_{\ell,m,B} = 0. \tag{20}$$

Appendix B gives more explicit formulations of the operators $\mathbf{K}$ and $\mathbf{G}$, using index notation on the sphere.

3.1.2 Flat-sky formulation

The flat-sky limits (corresponding to large ℓs) of equations (18)–(19) are (see Appendix C)

$$\mathbf{J} \approx (\mathbf{1}, \mathbf{0}) \quad \text{and} \quad \mathbf{Y} \approx \exp(i\boldsymbol{\ell}\cdot\hat{\boldsymbol{n}}), \tag{21}$$

while the parity eigenstates read locally, in the fixed copolar basis e_x, e_y:

$$\begin{aligned}\mathbf{W}^+ &\approx \cos(2\phi_\ell)\exp(i\,\boldsymbol{\ell}\cdot\hat{\boldsymbol{n}}) = -\frac{l_x^2-l_y^2}{l_x^2+l_y^2}\exp(i\,\boldsymbol{\ell}\cdot\hat{\boldsymbol{n}}),\\ \mathbf{W}^- &\approx i\,\sin(2\phi_\ell)\exp(i\,\boldsymbol{\ell}\cdot\hat{\boldsymbol{n}}) = -i\,\frac{2\,l_x\,l_y}{l_x^2+l_y^2}\exp(i\,\boldsymbol{\ell}\cdot\hat{\boldsymbol{n}}).\end{aligned} \tag{22}$$

In this limit, the *unknowns*, $\boldsymbol{a}$, represent the Fourier coefficients of the convergence field, κ. Note that our definition of γ and κ warrants that they are consistent with the lens equation on the tangent plane – solving for κ in equation (18) and plugging the solution into equations (22) – which reads locally in real space:

$$\nabla^2\kappa(\hat{\boldsymbol{n}}) = \left(\partial_x^2 - \partial_y^2\right)\gamma_1(\hat{\boldsymbol{n}}) + 2\,\partial_x\partial_y\gamma_2(\hat{\boldsymbol{n}}), \tag{23}$$

where $\gamma_1(x, y)$ and $\gamma_2(x, y)$ are the two components of the E and B modes of the shear field. Also note that thanks to equation (20) the recovered map will *not* have B modes by construction. It can nevertheless be checked that the amplitude of the B modes in the residuals is small compared to the amplitude of the signal in the E modes (see Section 4.2.3).

3.1.3 Cost function

The considered problem can be stated as recovering $\boldsymbol{a}$ given the data $\boldsymbol{g}$ according to the model in equation (17). In the same way as done for deblurring the images (Section 2), finding the solution of this inverse problem in the MAP (Pichon & Thiébaut 1998; Thiébaut 2005) sense involves minimizing a two-term cost function:

$$\mathcal{Q}(\boldsymbol{a}) = \mathcal{L}(\boldsymbol{a}) + \mu\,\mathcal{R}(\boldsymbol{a}), \tag{24}$$

with respect to the parameters $\boldsymbol{a}$. In the right-hand side of equation (24), the term $\mathcal{L}(\boldsymbol{a})$ enforces agreement of the model with the data, whereas $\mathcal{R}(\boldsymbol{a})$ is a regularization term used to enforce our prior knowledge about the sought fields, and $\mu \geq 0$ is a Lagrange multiplier used to tune the relative importance of the prior with respect to the data.

For errors with a centred Gaussian distribution, the likelihood term writes

$$\mathcal{L}(\boldsymbol{a}) = \sum_{j,k} W_{j_1,k_1,j_2,k_2}\, e_{k_1}(\hat{\boldsymbol{n}}_{j_1})\, e_{k_2}(\hat{\boldsymbol{n}}_{j_2}),$$

with $e_k(\hat{\boldsymbol{n}}_j) = g_k(\hat{\boldsymbol{n}}_j) - \gamma_k(\hat{\boldsymbol{n}}_j)/[1-\kappa(\hat{\boldsymbol{n}})]$ and $\mathbf{W} = \mathbf{C}^{-1}$ with $C_{j_1,k_1,j_2,k_2} = \langle e_{k_1}(\hat{\boldsymbol{n}}_{j_1})\, e_{k_2}(\hat{\boldsymbol{n}}_{j_2})\rangle$. If the errors are further uncorrelated, the likelihood simplifies to

$$\mathcal{L}(\boldsymbol{a}) = \sum_{j,k} w_{j,k}\left[g_k(\hat{\boldsymbol{n}}_j) - \frac{\gamma_k(\hat{\boldsymbol{n}}_j)}{1-\kappa(\hat{\boldsymbol{n}}_j)}\right]^2, \tag{25}$$

where the sum is carried over the index j of the sampled sky directions $\hat{\boldsymbol{n}}_j$ (so-called sky *pixels*) and index k of the two components of, say, the Q and U polarization fields, respectively (see Appendix B for an explicit formulation with all the relevant indices), and the weights are related to the variance of the noise:

$$w_{j,k} = \mathrm{Var}(e_k(\hat{\boldsymbol{n}}_j))^{-1}. \tag{26}$$

This allows us to account for non-uniform noise on the sky and also cuts (the galaxy, bright stars, etc.) for which the variance can

be considered as infinite and thus the corresponding weights set to zero. Note that setting the weights in this statistically consistent way yields no such biases as those which would result from interpolation or inpainting methods used to replace missing data [Pires et al. 2009; see also Abrial et al. (2008) for such implementation in the context of CMB experiments]. For this recovery problem, our prior is that the field κ must be as smooth as possible in the limit that the model remains compatible with observables within the error bars, that is equation (17) must be valid. To that end, the regularization is written as a penalty based on the second-order spatial derivatives (Laplacian) $\nabla^2\kappa$ of the field κ:

$$\mathcal{R}(\boldsymbol{a}) = \|\nabla^2\boldsymbol{\kappa}\| \,. \tag{27}$$

Equation (B12) in Appendix B gives the expression of $\nabla^2\kappa$ as a function of the unknown $\boldsymbol{a}$. In order to enforce smoothness while preserving some sharp features in the κ map, quadratic and non-quadratic norms of the Laplacian have been considered for the regularization (see Appendix B).

3.2 Generating the virtual data set

Let us first describe in turn the simulation used to model the full-sky κ map, and the generation of the corresponding map.

3.2.1 The simulation

The HORIZON-4Π (Prunet et al. 2008; Teyssier et al. 2009) simulation was used, a Λ cold dark matter simulation using the *WMAP* 3 cosmogony with a box size of $2h^{-1}$ Gpc on a grid of 4096^3 cells. 70 billion particles were evolved using the Particle Mesh scheme of the RAMSES code (Teyssier 2002) on an adaptively refined grid (AMR) with about 140 billions cells, reaching a formal resolution of 262 144 cells in each direction (roughly 7 kpc h^{-1} comoving). The simulation covers a sufficiently large volume to compute a full-sky convergence map, while resolving Milky Way size haloes with more than 100 particles, and exploring small scales deeply into the non-linear regime. The dark matter distribution in the simulation was integrated in a light cone out to redshift 1, around an observer located at the centre of the simulation box.

3.2.2 Mock data

This light cone was then used to calculate the corresponding full-sky lensing convergence field, which is mapped using the HEALPIX pixelization scheme with a pixel resolution of $\Delta\theta \simeq 0.74\,\mathrm{arcmin}^2 (n_{\rm side} = 4096)$. Specifically, the convergence $\kappa(\hat{\boldsymbol{n}})$ at the sky coordinate $\hat{\boldsymbol{n}}$ is computed from the density contrast, $\delta(\boldsymbol{x}, z)$, in the Born approximation using

$$\kappa(\hat{\boldsymbol{n}}) = \frac{3}{2}\Omega_{\rm m}\int_0^{z_{\rm s}} \frac{{\rm d}z}{E(z)}\frac{\mathcal{D}(z)\mathcal{D}(z,z_{\rm s})}{\mathcal{D}(z_{\rm s})}\frac{1}{a(z)}\delta\left(\frac{c}{H_0}\mathcal{D}(z)\hat{\boldsymbol{n}}, z\right), \tag{28}$$

which is valid for sources at a single redshift $z_{\rm s} = 1$, and $\mathcal{D}(z) = H_0\,\chi(z)/c$ is the adimensional comoving radial coordinate, hence ${\rm d}\mathcal{D} = {\rm d}z/E(z)$. The detailed procedure to construct such maps from the simulation using equation (28) is described in Appendices D1 (choosing the sampling strategy) and D2 and in Teyssier et al. (2009). In practice, a set of degraded maps of κ was generated from the full resolution, $n_{\rm side} = 4096$ down to $n_{\rm side} = 128$ in powers of 2, together with the corresponding masks (see Fig. 5). Different levels of noise (corresponding to $700 \leq \ell_{\rm cut} < 2500$) and maps with/without Galactic masks are considered. The corresponding simulations are labelled as ${}_{n_{\rm nside}}S^{\ell_{\rm cut}}_{\rm FS/GC}$. Cartesian maps are also used, labelled as ${}_{n_{\rm pixel}}C^{\rm S/N}_{\rm NL/lin}$ corresponding to Cartesian sections of the full-sky maps, where for commodity, the experiments involving high resolution were calibrated. Here, the flag NL/lin refers to whether or not the non-linear model is accounted for.

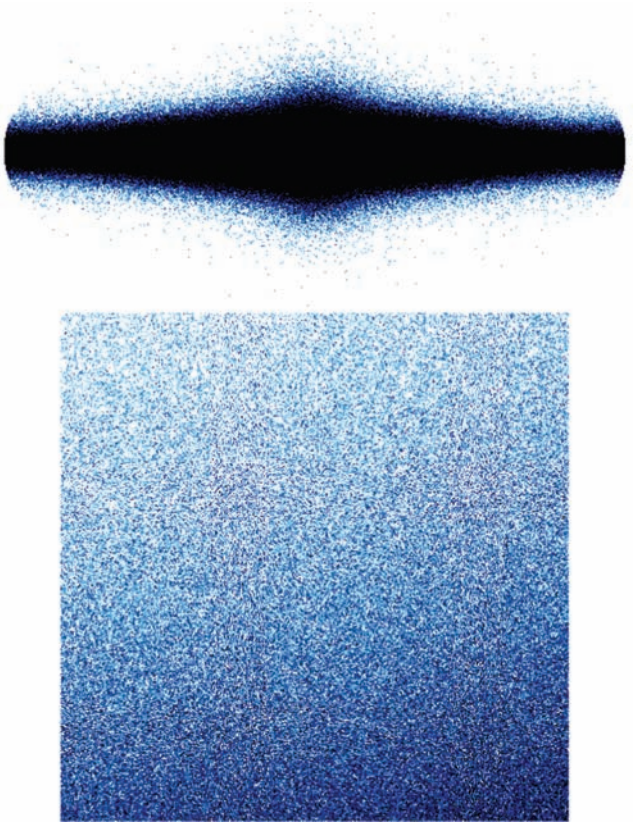

Figure 5. Top panel: full-sky view of the mask; bottom panel: a zoom at coordinate $(l, b) = (30^\circ, 30^\circ)$ showing the distribution of stellar cuts. This cut corresponds to the inner central region of the reconstruction shown in Fig. 12.

3.2.3 Penalty weight

In this paper, the weight of the penalty, μ, in equation (24) is chosen so that the ℓ_2 cut-off corresponds to the scale, $\ell_{\rm crit}$, at the intersection of the signal and the noise power spectra (see e.g. Fig. 6). Specifically

$$\mu \propto 1/\ell^2_{\rm crit}\,.$$

In a more realistic situation, when the power spectrum of the signal is unknown, GCV could be used to find this scale. When $\ell_1 - \ell_2$ penalty is implemented (see Section 2.1.4), the ℓ_1 parameter ϵ entering equation (16) is chosen so that it cuts off the tail of the PDF of the Laplacian of the recovered field at the 3σ level.

3.3 Optimization and performance

Let us now turn to the optimization procedure and the performance of the algorithm.

3.3.1 Optimization

Recall that the procedure assumes here a sampling strategy, since the noisy **g** field is given on a pixelization of the sphere. To solve the

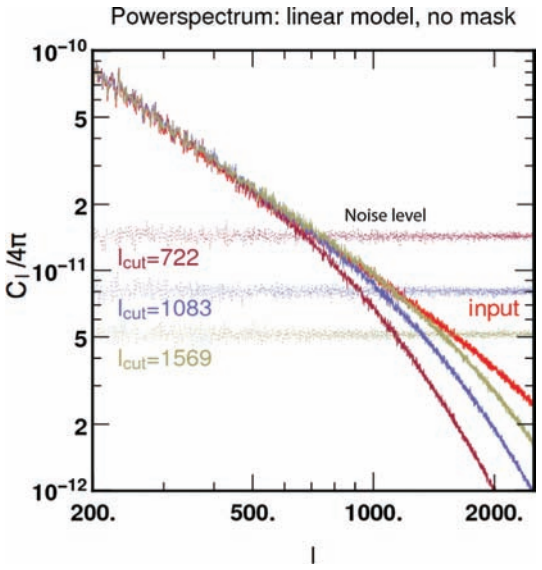

Figure 6. A zoom on the power spectra of the three reconstructions of ${}_{2048}S_{\rm FS}^{\ell_{\rm cut}}$ for $\ell_{\rm cut} = 722$, 1083 and 1569 (respectively 24, 44 and 78 ngal/□ arcmin), together with the power spectra of the noise. Note that the level of smoothing decreases with increasing S/N, in parallel to the bias in the corresponding power spectrum.

optimization problem, we used the algorithm VMLM from OPTIMPACK (Thiébaut 2002) which only involves computing the objective function $\mathcal{Q}(\boldsymbol{a})$ and its partial derivative with respect to the parameters $\boldsymbol{a}$. VMLM is an unconstrained version of VMLMB which has been used for the deblurring problem and which is described in some details in Section 2.1. The optimization of equation (24) is carried out by computing in turn equation (18) and equations (28) and (B7) using HEALPIX (Górski et al. 1999) in OPENMP or MPI.

3.3.2 Overall performance

Each back and forth transform takes, respectively, 0.1, 0.5, 2, 8, 32 and 128 s on an octo OPTERON for $n_{\rm side}$ equal to 128, 256, 512, 1024, 2048 and 4096 (see Table 2). The linearized problem without mask converges typically in a dozen iterations (which typically only involve a back and forth transform, unless the convergence is poor). The linearized mask problem takes a few hundred iterations (see Table 3), and so does the non-linear problem (or the linearized problem with a non-linear $\ell_1 - \ell_2$ penalty function).

Table 2. The performance of the optimization of the linearized inversion problem ${}_{n_{\rm side}}S_{\rm FS}^{\ell_{\rm cut}}$ as a function of $n_{\rm side}$ for an octo opteron in OPENMP.

$n_{\rm side}$	128	256	512	1024	2048
Time for one step (s)	0.13	0.59	2.13	8.48	34.3
Number of steps (s)	13	12	9	13	24
Total time (s)	2.6	10.4	33.4	171.1	1129.3

Table 3. Same as Table 2 with Galactic masks.

$n_{\rm side}$	128	256	512	1024	2048
Time for one step (s)	0.121	0.121	0.502	1.88	8.53
Number of steps (s)	252	313	315	377	325
Total time (s)	40.1	50.3	200	989	3340

4 VALIDATION AND POST-ANALYSIS

Let us illustrate on a sequence of statistical tests several crucial features of the ASKI map-making algorithm: its ability to fill gaps, its ability to preserve the geometry and sharpness of clusters, and maintain the gravitational nature of the signal in the presence of masks, and the freedom to choose strong/weak prior on the two-point correlation. These properties are important in various contexts of the weak lensing studies, such as the estimation of cosmological parameters, the physics of clusters, the interpretation of tomographic data from upcoming surveys, constraining the dark energy equation of state through the redshift evolution of statistical and topological tracers. We chose a selection of statistical tests that are sensitive to different aspects of map-making.

4.1 One-point statistics

4.1.1 Cluster counts

One of the main assets of high-resolution full-sky lensing maps is to probe multiple scales: it then becomes possible to sample the non-linear transition scale and, for example, study the shape of clusters. Fig. 7 illustrates this feature while displaying the result of the inversion with ℓ_2 and $\ell_1 - \ell_2$ penalties. For this experiment, a Cartesian subset at galactic coordinates $(l, b) = (0^\circ, 0^\circ)$ was extracted. The corresponding non-linear shear field $\mathbf{g}$ was generated via Fourier transform, and noised with a white additive noise of S/N of 1. This set was then inverted while assuming ℓ_2 (bottom right) and $\ell_1 - \ell_2$ (top right) penalties. The choice for the two penalty weights, μ and ϵ, was made on the basis of least-square residual in the inverted κ maps. The improvement of $\ell_1 - \ell_2$ over ℓ_2 penalty is significant. This statement is made more quantitative in Fig. 8

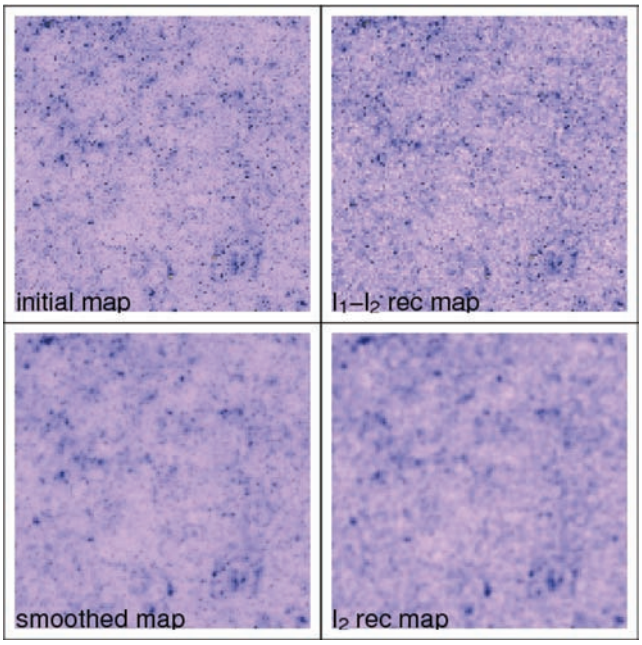

Figure 7. Top-left panel: a zoom of the original map at coordinates $(l, b) = (0^\circ, 0^\circ)$; top-right panel: reconstruction with $\ell_1 - \ell_2$ penalty using the non-linear model. Bottom-left panel: input map smoothed at an FWHM of 1.5 pixels. Bottom-right panel: reconstruction with ℓ_2 penalty using the non-linear model. The colour table is linear. The edge-preserving penalty appears qualitatively to preserve much better the amplitude and the number of high peaks in the κ map, as shown quantitatively in Fig. 8.

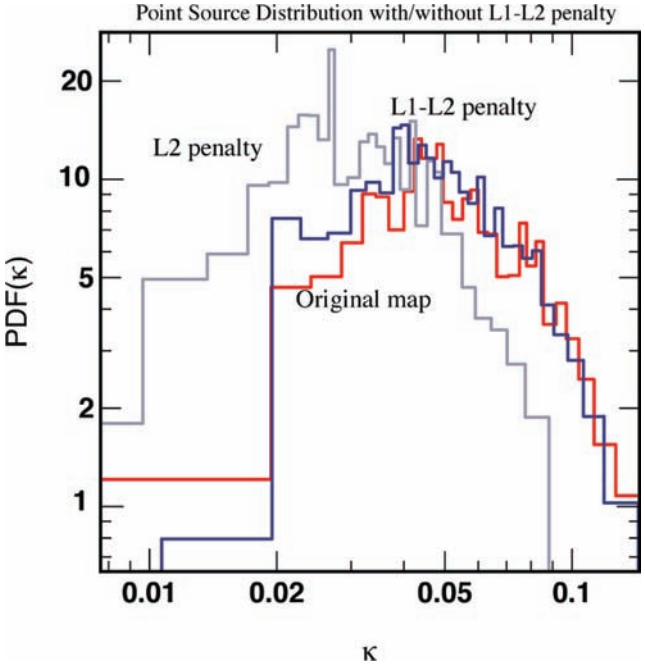

Figure 8. The PDF of point source as defined in Section 4.3.1 corresponding to the maps of Fig. 7 recovered with $\ell_1 - \ell_2$ penalty, and ℓ_2 penalty, respectively. The improvement with an edge-preserving penalty is significant.

which displays the PDF of the peaks within that image for the initial map (top-left panel of Fig. 7) computed following the peak patch prescription described in Section 4.3.1. The agreement between the input and the recovered distribution is significantly enhanced by the optimal $\ell_1 - \ell_2$ (top right) penalty.

4.1.2 Skewness and Kurtosis

The simplest statistics to explore the non-linear transition are the skewness, S_3, and the kurtosis, S_4, of the PDF of the recovered maps. Furthermore, it has been shown that these parameters provide a powerful tool to measure the underlying cosmological parameters (Bernardeau, van Waerbeke & Mellier 1997; Takada & Jain 2002, 2004). Fig. 9 displays the evolution of these numbers as a function of scale in the initial and recovered maps, with and without galactic masking. The top hat filter used here is of width $[2^i, 2^{i+1}]$, while the harmonic number of each band is the mean of its boundary: $\bar{i} = (2^i + 2^{i+1})/2$. The recovery of skewness and kurtosis is good in the case of unmasked data. Of course, it degrades with the scale as we reach $\ell_{\rm cut}$. Using the reconstructed map is not the optimal way of measuring the three- and four-point functions at small scale. However, an optimal dedicated estimator can be built upon the same regularization technique. The masked case is not as good. There, a dedicated estimator, acting only on small, clean, pieces of the sky will probably yield better results.

4.1.3 Accounting for a non-linear model

Fig. 10 shows the effect of accounting for the non-linearity in equation (17). Here, a set of Cartesian simulations is used ${}_{256}C^{1}_{\rm NL/lin}$. This map represents (a 100 times) the difference between the recovered map while accounting for $1 - \kappa$ in equation (17) in the inversion, and the recovered map while neglecting this factor. The difference is small in amplitude, but shows as expected the strongest bias near the clusters and the filaments, where κ is largest. The bottom panel represents the corresponding relative power spectrum, $C_\ell[{\rm NL} - {\rm lin}]/C_\ell[{\rm input}]$, as a function of ℓ. Again, the larger discrepancy occurs at higher ℓ, corresponding to the sharp peaks at the positions of the clusters. Hence, the non-linearity should be accounted for in the model if the shape of the cluster is an issue (see also White 2005; Dodelson & Zhang 2005; Shapiro 2009). For all practical purposes, we have therefore demonstrated that at scales below $\ell_{\rm max} < 4096$, solving the linearized problem is de facto equivalent to the general non-linear problem when κ is neglected at the denominator in equation (17).

4.2 Two-point statistics

Since ASKI was constructed to provide the optimal *map* given the measured shear, we do not expect that it will yield *the* optimal estimator for non-linear functions of these maps, such as the power spectrum, bispectrum, etc. Nevertheless, it is of interest to compare the two-point statistics of input and recovered maps to see how

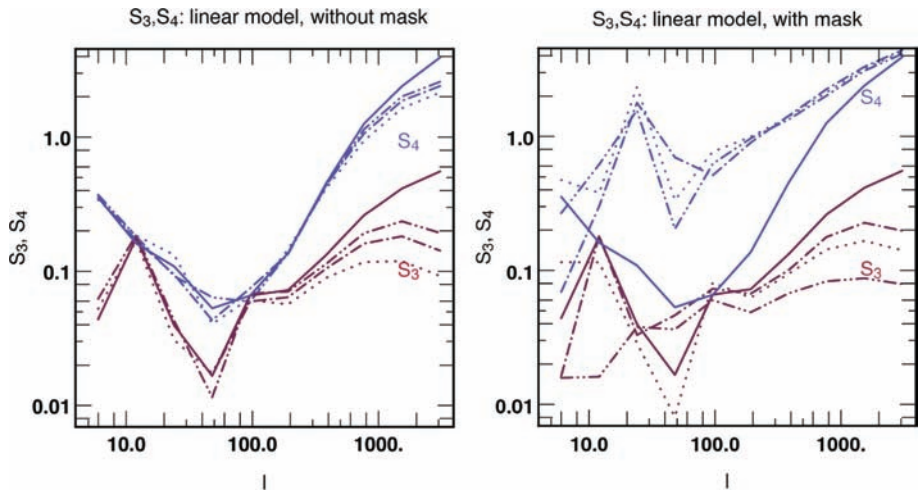

Figure 9. Left-hand panel: skewness, S_3 and kurtosis, S_4 as a function of scale (using sharp top hat filtering) for the model (plane line) and the recovered κ maps of a simulation ${}_{2048}S^{\ell_{\rm cut}}_{\rm FS}$ (dotted, dot–dashed, dot-dot–dashed line for $\ell_{\rm cut} = 722$, 1083 and 1569, respectively, 24, 44 and 78 ngal/$\square$ arcmin); right-hand panel: same as top panel, but for the ${}_{2048}S^{\ell_{\rm cut}}_{\rm GC}$ set. Note that the kurtosis of the cut is significantly different at small ℓ.

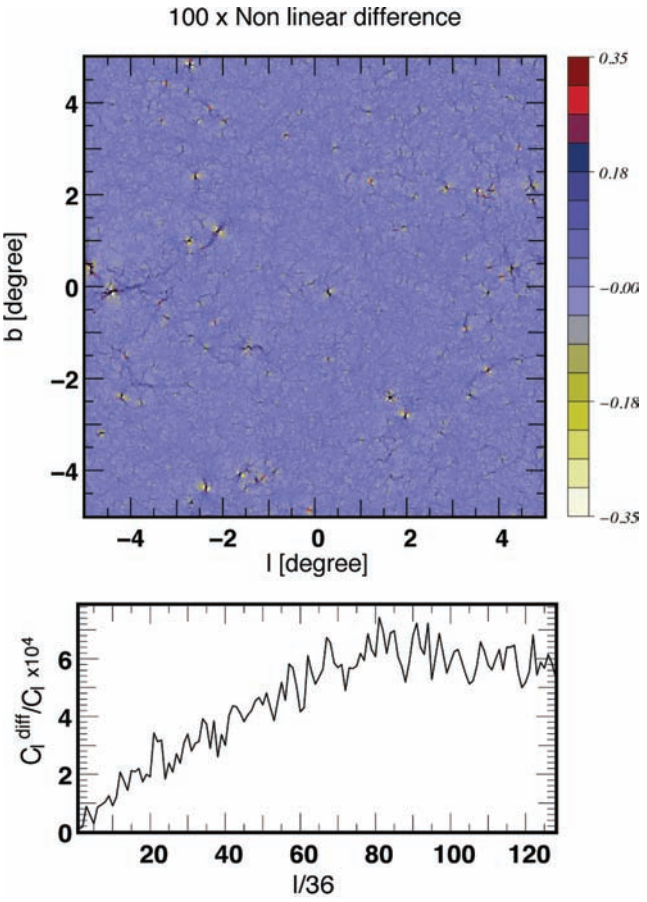

Figure 10. Top panel: a map of the 100 times difference between the recovered map with the non-linear model and $\ell_1 - \ell_2$ penalty, and the recovered map without accounting for the non-linearity. As expected the difference is largest at high frequencies near the cluster and along the filaments. Bottom panel: the power spectrum of the relative difference as a function of ℓ.

ASKI deals with masks and how it affects the occurrence of spurious B modes.

4.2.1 *Optimal Wiener filtering*

Throughout this paper, the prior $C_\ell \equiv \ell^{-1}(\ell + 1)^{-1}$ ('Laplacian prior') is used in equation (B.11). Let us briefly investigate how a customized (Wiener) prior for C_ℓ changes the reconstruction at small scales. Fig. 11 shows that the corresponding power spectra of the reconstructed κ maps, as expected, differ mostly for scales where the S/N is smaller than 1. However, when the smoothing (Laplacian) prior amplitude (see equation 27) is tuned to minimize the reconstruction error as in the figure, the power spectra of the reconstructed maps for the two different priors (Laplacian and Wiener) are quite similar.

It is interesting to note that the power of the reconstructed map with the Wiener prior (light brown line in Fig. 11) is systematically biased low as compared to the input power spectrum. This reflects the fact that an optimal (minimum variance) estimation of the power spectrum *is not equivalent to* a power-spectrum estimation on an optimal (minimum variance) reconstructed map. However, in the simple case where we have noisy data without masks, the bias of the power spectrum of the minimum variance map reconstruction is known, it is simply given by $C_\ell/(C_\ell + N_\ell)$ where C_ℓ is the power

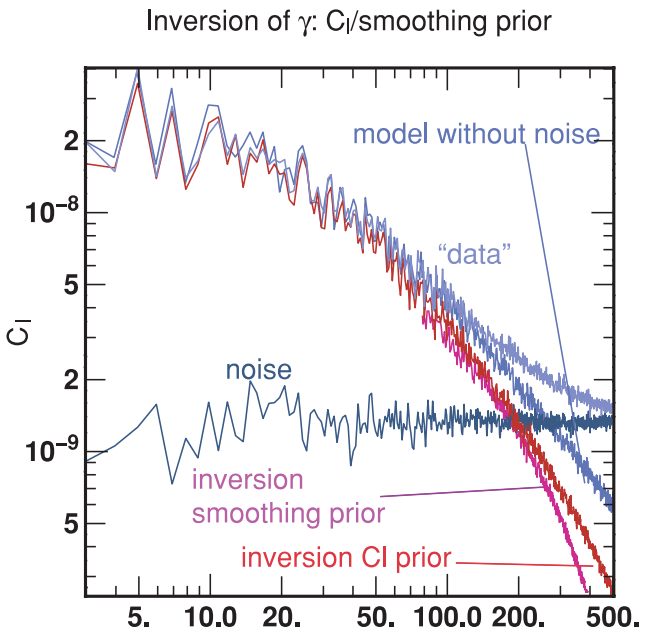

Figure 11. The power spectrum of the input and recovered κ [with smoothing and C_ℓ prior, see equation (B8)] as a function of ℓ, together with the power spectrum of the noise and the noisy equivalent κ using a simulation, ${}_{512}S^{224}_{\rm FS}$ (i.e. ngal/$\square$ arcmin $= 5$). Note that the recovered power spectrum departs from the power spectrum of the input field roughly at the cut-off frequency when a quadratic smoothing penalty is applied.

spectrum of the underlying κ map (without noise), and N_ℓ is the noise power spectrum in 'kappa' space, which is given approximately in our case by $\sigma^2\Omega_{\rm pix}$, where σ^2 is the noise variance per pixel in the shear field and $\Omega_{\rm pix}$ is the solid angle of a pixel.

In Fig. 11, the noise level is shown by the horizontal dark blue line. One can see in particular in the figure that when the model power spectrum (without noise) crosses the noise power spectrum, the power spectrum of the minimum variance map (golden line) is lower than that of the model by a factor of 2, as expected from the considerations above, even in the presence of masks. Thus, an approximate, but simple, way to get an unbiased estimate of the kappa power spectrum is to correct the minimum variance map power spectrum by the ratio $C_\ell/(C_\ell + N_\ell)$. Note, however, that a true minimum variance power-spectrum estimation of the kappa field is not the aim of the present method (see e.g. Pen 2003 for the flat-sky case).

Nevertheless, elsewhere in this paper, a smoothing prior which is not customized to the specific problem is preferred.

4.2.2 *Filling gaps within masks*

Let us first compare visually the recovered map to the input map. Fig. 12 illustrates a feature of the penalized reconstruction: it interpolates quite well and provides means to fill the gaps corresponding to the galactic cuts. For a more quantitative comparison, we also plot in this figure the ridges of both fields (using the skeleton, see below), which match very well up to the very edge of the mask. The smoothing penalty also induces a level of extrapolation, best seen in the residuals (see Fig. 13). The masking (or more generally, non-uniform weights, w_i) nevertheless biases the reconstructed map, as seen in Figs 9 and 14. Note finally that when masks are accounted for, it is straightforward to correct for them when computing the

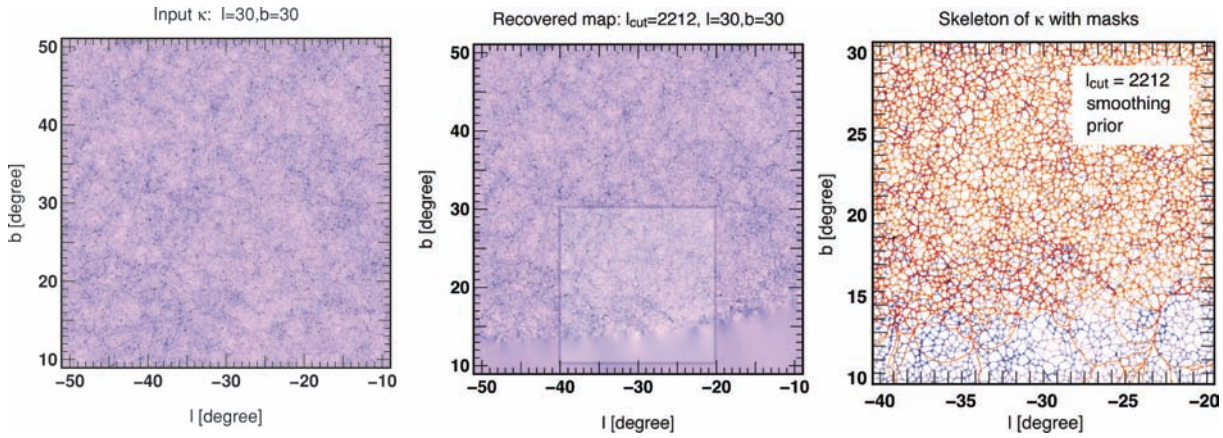

Figure 12. Left-hand panel: the initial κ map in the region with bright stars masking shown in Fig. 5 at coordinates $(l, b) = (30^\circ, 30^\circ)$; middle panel: the corresponding recovered κ map of a simulation ${}_{2048}\,S_{\rm GC}^{2212}$. Note that the gaps have been nicely filled up to the very edge of the mask; right-hand panel: the corresponding two skeletons (colour coded by κ in purple: input skeleton; in orange: recovered skeleton) for the inner region (marked as a square on the middle panel), when masking is present. Note the clear gradient away from the mask in the quality of the match between the two skeletons; recall that most of this field is partially shielded by stars, as seen in Fig. 5.

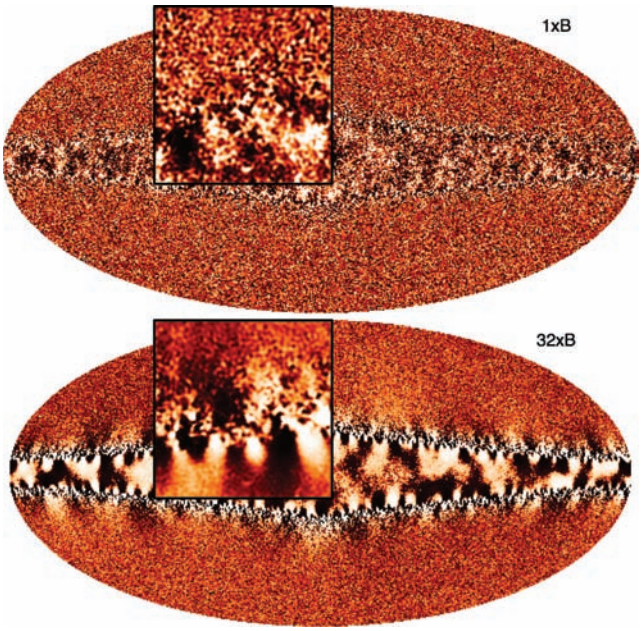

Figure 13. An example of full-sky leak of the B modes when masks are accounted for; top panel: the residuals corresponding to $\sigma_{\rm B} = \sigma + \sigma$; the inner box corresponds to a zoom near the edge of the galactic cut at $(b, l) = (30, 20)$; bottom panel: same residual and box for $\sigma_B = \sigma + 32\sigma$. Note that for the latter case, the extent of the leakage is much larger and coherent.

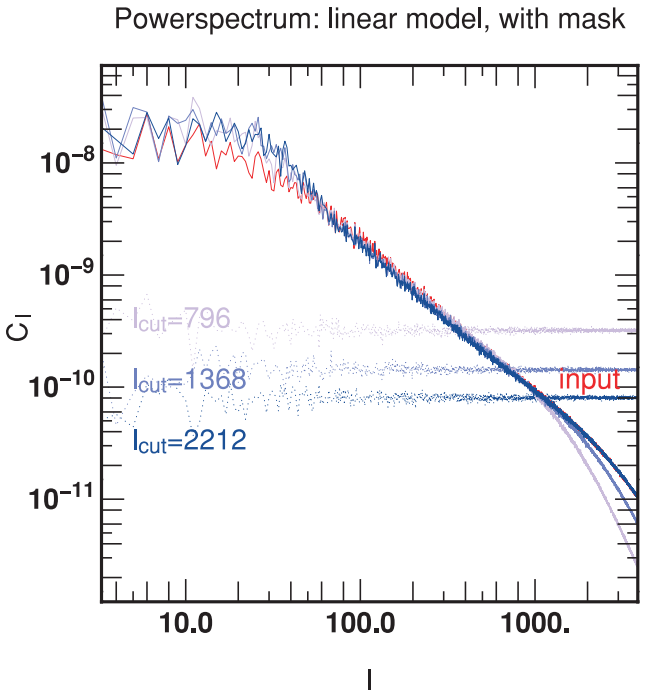

Figure 14. The power spectra of three high-resolution reconstructions corresponding to Fig. 12 for $\ell_{\rm cut} = 796$, 1368 and 2212 (respectively 28, 63 and 130 ngal/$\square$ arcmin corresponding to a low-, intermediate- and high-end values) together with the power spectra of the noise. Note that the recovered power spectrum has extra power at large scales and less power at intermediate scales, an artefact of the mask which can be corrected for by accounting for the prior knowledge of the auto-correlation of the mask.

power spectrum as the harmonic transform of the autocorrelation, which in turn is derived by correcting for the autocorrelation of the masks (see Szapudi, Prunet & Colombi 2001; Hivon et al. 2002; Chon et al. 2004 for details). When seeking the three-point correlations, one could also proceed accordingly, and divide by the three-point correlation of the mask. Indeed, a three-point-reduced correlation is simply one plus the excess probability of finding triplets, which in turn is computed by counting the number of found triplets and dividing by the expected number of such triplets given the shape of the mask (Chen & Szapudi 2005). This also applies if the mask is grey.

4.2.3 *Residual B modes*

Let us investigate the effect of leaking of B modes with the following experiment: the noise in the transform of the B channel is boosted by some fixed amount over a map which has Galactic cuts. This corresponds to the case where the B is significantly larger than the noise, yet uncorrelated with the E mode, corresponding to, for example, a systematic bias in the ellipticity extraction. It is expected that, due to masks, this B mode will leak in E. An example of such

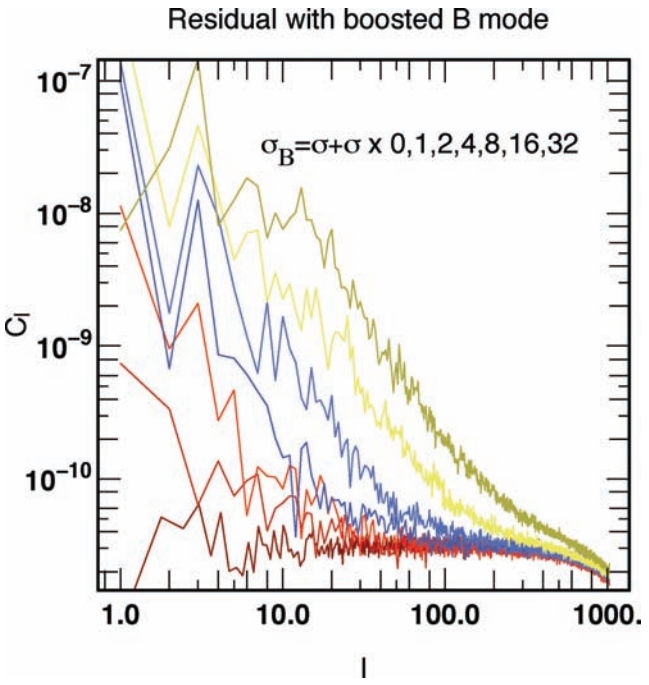

Figure 15. Power spectrum of the mask weighted residual error on κ as a function of the harmonic number, ℓ. The different curves correspond to boost the B modes of increasing relative strength. The low-order modes are polluted by leaks from the masks (see also Fig. 13); here $l_{\rm cut} = 752$ (60 ngal/□ arcmin).

leak, for B modes as large or up to 32 times larger than the noise, is shown in Fig. 13. The power spectrum of the residuals in the corresponding κ map is computed while masking in the residual the exact regions corresponding to the cuts. When this boost is zero (bottom curve in Fig. 15), the power spectrum of these residuals is flat and corresponds to the noise power spectrum. In contrast, the stronger the boost the larger the scale below which this power spectrum is coloured. Note that it was checked that, as expected, these coherent residuals disappear completely if the galactic cuts are ignored. It would also be interesting to compare the distribution of the shape of dark matter in input/recovered clusters.

Finally, note that Appendix C2 discusses briefly the effect of noise in power-spectrum estimation.

4.3 Alternative statistics: critical sets

Let us close this section with a quantitative comparison of the input and the recovered map using more exotic probes to estimate the quality of the reconstruction, and the prospect it offers for dark energy measurements. Indeed, the predictions of the perturbative hierarchical clustering model are often given through the hierarchy of the differences between the moments to their Gaussian limit. Yet higher order moments are generally difficult to test directly in real-life observations, due to their sensitivity to very rare events. As argued in Pogosyan, Gay & Pichon (2009a) the geometrical analysis of the critical sets in the field (extrema counts, Genus, critical lines, etc.) may provide more robust measures of non-Gaussianity, and is becoming elsewhere an active field of investigation (Park et al. 2005; Gott et al. 2009).

4.3.1 Peak patch counts and area

Even though many tools are available to identify peaks within the reconstructed map, let us validate here our reconstruction using a segmentation of both the initial and the recovered maps using peak patches on the sphere, which are a segmentation of the map based on the attraction patches of the κ map when following its gradient (see Sousbie, Colombi & Pichon 2009). Within each peak patch (see Fig. B1), the brightest pixel is assigned a mass corresponding to the enclosed mass within the peak patch. This quantity is gravitationally motivated (as the patch corresponds to the attraction region of the cluster) and is both robust (as the geometry of the patch only depends on the imposed smoothing length, which in turn is fixed by the resolution of the survey) and sensitive to small features in the map; it is therefore a good indicator of the quality of the reconstruction. Fig. 16 (left-hand panel) displays the corresponding PDFs before and after reconstruction. As expected, the recovered point source PDF has a shifted mode and is less skewed

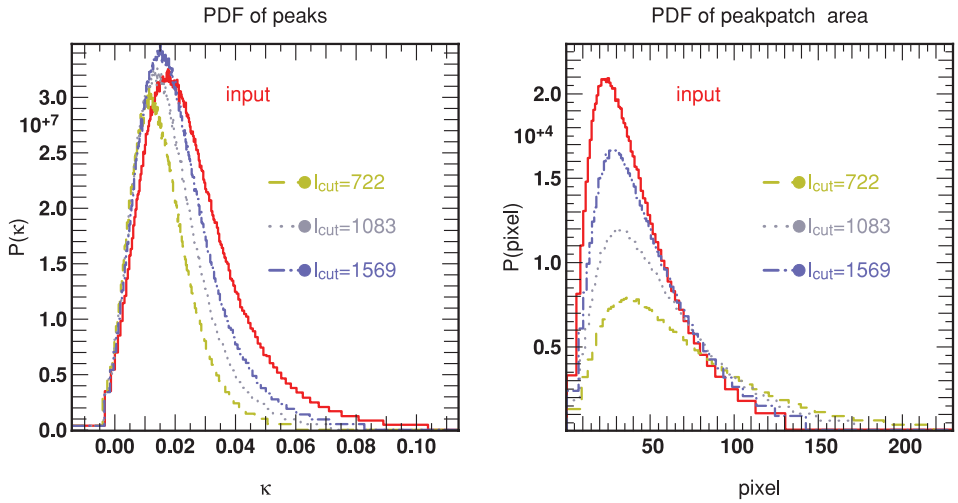

Figure 16. Left-hand panel: the PDFs of $\kappa_{\rm max}$ at point sources before and after reconstruction of a set of simulations ${}_{2048}S_{\rm FS}^{\ell_{\rm cut}}$ [dashed, dotted, dot–dashed line for $\ell_{\rm cut} = 722$, 1083 and 1569 (respectively 24, 44 and 78 ngal/□ arcmin)]; right-hand panel: the PDFs of the area of peak patches (see Fig. B1) before and after reconstruction for the same set of simulations. Note that, as expected, the recovered distribution of peaks is less skewed than the original, whereas conversely, the PDF of the area of the peak patches for the low S/N reconstruction is more skewed towards larger patches.

than the original distribution. This trend decreases with increasing S/N. For realistic galaxy counts of 40 ngal/□ arcmin, the agreement between the input and the recovered PDF is fairly good, and the corresponding residual bias can be modelled (as the reconstruction is essentially a smoothing of the underlying map). This could lead to interesting constraints on Ω_m and $D(z)$ when used in conjunction with weak lensing tomography in order to probe its redshift evolution. The right-hand panel of Fig. 16 focuses on a different quantity, the area of the patches, which when compared to the area of the corresponding void patches could also be used as a measure of the gravitationally induced non-Gaussianities, together with their shape (higher moments of κ within a patch). Again, the reconstruction seems to recover this distribution well enough to suggest that such a tool could be used in the future to study the cosmic evolution of the projected web.

4.3.2 *Topology & geometry: critical lines*

Let us now compare the shape of the recovered map to the initial map from the point of view of its critical lines. For this purpose, let us use here the skeleton as a geometric probe (Novikov, Colombi & Doré 2006; Sousbie et al. 2008b). It is defined in 2D as the boundary of the void patches, which in turn are a segmentation of the map based on the valleys of the κ map [corresponding to the peak patches (defined above) of *minus* the field]. The skeleton of the initial field and the recovered fields for simulation ${}_{2048}S_{\rm GC}^{\ell_{\rm cut}}$ is computed, and represented in Fig. 12. The recovered skeletons are qualitatively fairly close to the original skeleton, which demonstrates that the local topology and geometry of the field are well recovered. Let us make this comparison more quantitative. The differential length per unit area of the recovered field (*the set* ${}_{2048}S_{\rm FS}^{\ell_{\rm cut}}$ with $\ell_{\rm cut} = 722, 1083$ and 1569 *as labelled*)[9] over the initial κ map (thin line) as a function of density threshold is also shown in Fig. 17, while Fig. 18 shows the corresponding maps for similar runs, together with a map of the orientation of the **ssg** field. The agreement increases at larger density thresholds, which suggests that the topology of dense regions is well recovered.[10] The total length was shown (Sousbie et al. 2009) to trace well the underlying shape parameter of the power spectrum and has been used in 3D to constrain the dark matter content of the Universe (Sousbie et al. 2008a). As shown in Pogosyan et al. (2009b) this would work for 2D maps, and could therefore be used with κ maps such as those reconstructed via the present method. The redshift evolution of this differential count, when tomographic data are available, could complement, for example, Genus measurements as means of constraining the dark energy equation of state in a manner which could be more robust than direct cumulant estimation. Eventually, the skeleton could also be used to characterize the connectivity of clusters (i.e. the number of connected projected filaments), as it will also depend on the cosmic dark energy content of the universe (Pichon al. 2009).

This rapid review has shown that depending on the final objective (cosmological parameters, cross-correlation with other maps, etc.), a variety of estimators can be extracted from the recovered maps. ASKI was shown to perform rather well with respect to these estimators. Defining the best combination of these estimators – and the optimal penalty associated – will be one of the key topics for lensing research for the coming years.

[9] Note that ngal/□ arcmin $= 40(l_{\rm cut}/1000)^{1.5}$.

[10] In fact, the relative distance between the recovered and the input skeleton could also be used as an alternative to the differential length (see Caucci et al. 2008).

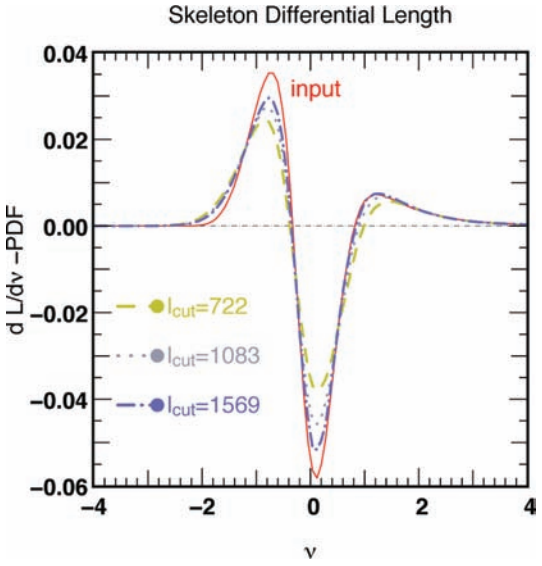

Figure 17. The input skeleton differential length (a tracer of Ω_m) with its recovered counterparts as a function of the normalized κ contrast, $\nu \equiv (\kappa - \bar{\kappa})/\sigma_\kappa$ for the set ${}_{2048}S_{\rm FS}^{\ell_{\rm cut}}$ with $\ell_{\rm cut} = 722$, 1083 and 1569 (respectively 24, 44 and 78 ngal/□ arcmin). Here, the PDF of the normalized κ contrast was subtracted to the differential length for clarity. As expected, the agreement is best at large convergence. This figure is complementary to Fig. 12 which shows that the *geometry* of the field is well preserved on average.

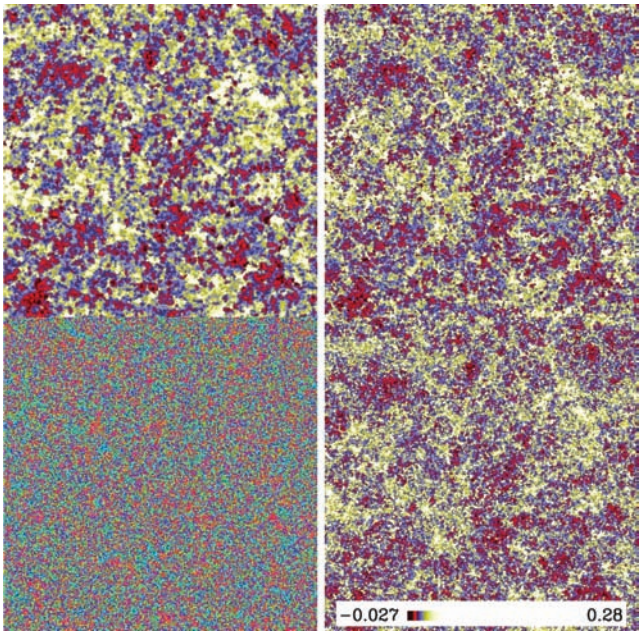

Figure 18. A zoom of the full-sky recovered κ maps of a simulation ${}_{2048}S_{\rm FS}^{\ell_{\rm cut}}$ with $\ell_{\rm cut} = 722$ (top-left panel) and 1569 (top-right panel) (respectively 24 and 78 ngal/□ arcmin) at coordinates $(\phi, \theta) = (0, 0)$ (the colour table corresponds to a histogram equalization); bottom-left panel: the corresponding data (the *hue* colour table codes the shear orientation); bottom-right panel: the corresponding underlying κ map.

5 CONCLUSION AND DISCUSSION

This paper sketched possible solutions to issues that a full-sky weak lensing pipeline will have to address, and presented an inverse method implementing the deblurring of the image and the map-making step.

Weak lensing surveys require measuring statistical distributions of the morphological parameters (ellipticity, orientation, ...) of a very large number of galaxies. This paper demonstrated that these parameters can be measured with a better accuracy and strongly reduced bias if the deep sky images are properly deblurred prior to the shape measurements. Using a *relative* figure of merit (the recovered SExtractor ellipticity) we have shown that this deblurring could in crowded fields improve more than tenfold the accuracy of the recovered ellipticities. This deblurring is critical in crowded regions, where the overlapping of stars and galaxies otherwise prevents accurate morphological estimation. Henceforth, dealing with such regions is important for a full-sky survey. Since such surveys will require the processing of a great number of large images, the calibration of these techniques is automated on the images themselves via cross-validation *after* the identification and removal of the stars within the field (see Fig. 1). In particular, the level of regularization, μ and the $\ell_1 - \ell_2$ threshold are *automatically* tuned in order to deal with the noise level and the dynamics of the raw images. The gap-filling interpolation feature of the inversion would apply even more efficiently in this regime than in the map reconstruction regime described in Section 3. The algorithm described here scales well since it only relies on DFTs: hence it could be applied to very large images such as those produced by modern surveys. aski uses the efficient variable metric limited memory algorithm optimPack, which allows both optimizations to scale to high resolutions. The deblurring is implemented on Cartesian maps as large as $16\,384^2$ pixels. GCV was shown to yield a quantitative threshold in order to accurately remove the point sources within the field, hence imposing the optimal level of smoothing for the galaxies only. In this paper, the focus was put on blurred 8-m ground-based observations, but the implementation for EUCLID-like space missions should be straightforward. The above-described improvements could clearly be reproduced if alternative state-of-the-art shear estimators were to be used (as compared by the SHear Testing Program).

This paper also demonstrated that optimization in the context of MAP provided a consistent framework for the optimal reconstruction of κ maps on the sphere. The main asset of the aski algorithm is that the penalty can be applied in model space, while the optimization iterates back and forth between data space and model space. This freedom allows it to deal simultaneously with masks (in data space) and edge-preserving penalties. Providing κ maps is critical in its own right, as it maps the dark matter distribution of our universe, and gives access to the underlying power spectrum on large scales. Such maps are also interesting when cross-correlated with other surveys (optical surveys, CMB maps, lensing reconstruction and distribution of SZ clusters from the Planck mission, redshift evolution of X-ray sources counts, etc.) in order to explore the evolution of the large-scale structure, and in the case of the surveys mapping the baryonic matter, to better understand biasing as a function of scale. Finally, though not optimally, it can be used to compute second and higher order statistics, and noticeably the three-point statistics, the Genus, cluster counts or the skeleton, which may constrain more efficiently the dark energy equation of state, as they are less sensitive to rare events. It should be stressed once more that while the reconstructed κ maps yield biased estimates of the power-spectrum and higher order statistics, the technique described in this paper can be adapted to build dedicated optimal estimators for each of those observables.

Section 4 demonstrated the quality and limitations of the reconstruction using various statistical tools on a full-sky simulation of $\mathbf{g}$ with resolutions of up to $12 \times 4096^2 = 201\,326\,592$ pixels thanks again to the efficiency of optimpack. In particular, it identified point sources of the fields, analyzed their PDF and showed that $\ell_1 - \ell_2$ penalty was critical at small scales. It also investigated the effect of leakage of B modes when Galactic cuts are present. It presented a method to probe the topology and geometry of a field on the sphere, the peak patches and the skeleton, and applied it to compare the recovered field to the initial field. Such tools allow us to quantify the differences between the two maps and act as an efficient source segmentation algorithm. Indeed, the degeneracy between the cosmological parameters ($\Omega_{\rm M}$, σ_8) is, for instance, best lifted with cluster counts. They may also turn out to be of importance when probing the dark energy equation of state as they are less sensitive to rare events. The Cartesian dual formulation of aski was also implemented and may prove useful for surveys where sky coverage is sufficiently small.

In short, aski accounts for the possible building blocks that a full-scale pipeline aiming at sampling the dark matter distribution over the whole sky should provide. Specifically, it allows for (i) automatically deblurring very large images using non-parametric self-calibrated edge-preserving $\ell_1 - \ell_2$ deconvolution with positivity; (ii) carrying the large *non-linear*inverse problem of reconstructing the convergence κ from the shear $\mathbf{g}$ using equation (17): the back and forth iterations between model and data are consistent with constraints in both spaces, and allow for an accurate recovery of cluster profiles and shapes; (iii) *non-uniform* weighting and masking: consistent with realistic Galactic cuts (and bright stars masking) and non-uniform sampling of the different regions of the sky, dealing transparently with the issue of the boundary; (iv) edge-preserving $\ell_1 - \ell_2$ penalty yielding quasi-point-like cluster reconstruction. Finally, (v) it introduced peak patches and the skeleton on the sphere, together with its statistics.

Possible improvements/investigation beyond the scope of this paper involve: (i) comparing the absolute gain in shear estimation using alternative tools to SExtractor (such as Massey et al. 2007) with more realistic galactic shapes; (ii) deblurring the images with a variable PSF within the field; (iii) building optimal estimators for the power spectrum C_ℓ^κ, or the asymmetry S_3 (a possible option would be to rely on perturbation theory, and invert the non-linear problem for both C_ℓ^κ and S_3); (iv) inverting for γ *and* κ *simultaneously* and checking a posteriori the amplitude of the B modes [an alternative to the model described in equation (19); the issue of unicity of the solution will be a challenge]; (v) carrying the deprojection while assuming prior knowledge of a complete distribution of source planes in equation (28) (the corresponding inverse problem remains linear, with an effective kernel which depends on the optical configuration and the distribution of galaxies as a function of redshift); (vi) moving away from the Born approximation, which involves solving Poisson's equation for each slice, and ray-tracing back to the source while solving for the lens equation though all the slices; (vii) implementing a more realistic noise modelling [which amounts to changing the cost function, equation (24)]; (viii) studying the shape of dark matter distribution in clusters and groups: typically this would also involve cross-correlating the corresponding distribution with the light at various wavelengths, (ix) defining the post-analysis which is most sensitive to dark energy, given the feature of the surveys to come and finally (x) propagating the analysis up to the cosmic figure of merit for the dark energy parameters.

ACKNOWLEDGMENTS

We thank Dmitry Pogosyan, Dominique Aubert, Eric Hivon, Martin Kilbinger and Yannick Mellier for comments and suggestions, the horizon-4Π team and the staff at the CCRT for their help

in producing the simulation, and D. Munro for freely distributing his YORICK programming language and OPENGL interface (available at http://yorick.sourceforge.net/). The galactic mask was provided to us by Adam Amara. This work was carried within the framework of the HORIZON project: http://www.projet-horizon.fr.

REFERENCES

Abrial P., Moudden Y., Starck J.-L., Fadili J., Delabrouille J., Nguyen M. K., 2008, Statistical Methodology, 5, 289
Aubert G., Kornprobst P., 2008, Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations (Applied Mathematical Sciences), 1st edn. Springer-Verlag, Berlin
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bartelmann M., Narayan R., Seitz S., Schneider P., 1996, ApJ, 464, L115
Benabed K., Scoccimarro R., 2006, A&A, 456, 421
Bernardeau F., van Waerbeke L., Mellier Y., 1997, A&A, 322, 1
Bernardeau F., Mellier Y., van Waerbeke L., 2002, A&A, 389, L28
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bradac M., Schneider P., Lombardi M., Erben T., 2005, A&A, 437, 39
Bridle S., Abdalla F., 2007, ApJ, 655, L1
Bridle S., Hobson M., Lasenby A., Saunders R., 1998, MNRAS, 299, 895
Cacciato M., Bartelmann M., Meneghetti M., Moscardini L., 2006, A&A, 458, 349
Caucci S., Colombi S., Pichon C., Rollinde E., Petitjean P., Sousbie T., 2008, MNRAS, 386, 211
Chen G., Szapudi I., 2005, ApJ, 635, 743
Chon G., Challinor A., Prunet S., Hivon E., Szapudi I., 2004, MNRAS, 350, 914
Crittenden R., Natarajan P., Pen U., Theuns T., 2002, ApJ, 568, 20
Dodelson S., Zhang P., 2005, Phys. Rev. D, 72, 083001
Erben T., Van Waerbeke L., Bertin E., Mellier Y., Schneider P., 2001, A&A, 366, 717
Fu L. et al., 2008, A&A, 479, 20
Girard D. A., 1989, Numr. Math., 56, 1
Golub G. H., Heath M., Wahba G., 1979, Technometrics, 21, 215
Górski K. M. et al., 1999, in Banday A. J., Sheth R. K., da Costa L. N., eds, Proc. MPA-ESO Cosmology Conf., Evolution of Large Scale Structure: From Recombination to Garching. ESO, Garching, p. 37
Gott J. R., Choi Y.-Y., Park C., Kim J., 2009, ApJ, 695, L45
Halkola A., Seitz S., Pannella M., 2006, MNRAS, 372, 1425
Heymans e. A., 2006, MNRAS, 368, 1323
Hirata C., Seljak U., 2004, Phys. Rev. D, 70, 063526
Hivon E., Górski K. M., Netterfield C. B., Crill B. P., Prunet S., Hansen F., 2002, ApJ, 567, 2
Högbom J. A., 1974, A&AS, 15, 417
Hu W., 2000, Phys. Rev. D, 62, 043007
Jee M. et al., 2007, ApJ, 661, 728
Kilbinger M., Schneider P., 2005, A&A, 442, 69
Kitching T., Heavens A., Taylor A., Brown M., Meisenheimer K., Wolf C., Gray M. E., Bacon D. J., 2006, MNRAS, 376, 771
Marshall P., Hobson M., Gull S., Bridle S., 2002, MNRAS, 335, 1037
Massey e. A., 2007, MNRAS, 376, 13
Massey R. et al., 2007, ApJS, 172, 239
Mugnier L. M., Fusco T., Conan J.-M., 2004, J. Opt. Soc. Am. A, 21, 1841
Nocedal J., Wright S. J., 2006, Numerical Optimization, 2nd edn. Springer-Verlag, Berlin
Novikov D., Colombi S., Doré O., 2006, MNRAS, 366, 1201
Park C. et al., 2005, ApJ, 633, 11
Peacock J. A., 1999, Cosmological Physics. Cambridge Univ. Press, Cambridge
Pen U.-L., 2003, MNRAS, 346, 619
Pichon C., Bernardeau F., 1999, A&A, 343, 663
Pichon C., Thiébaut E., 1998, MNRAS, 301, 419
Pichon C., Vergely J. L., Rollinde E., Colombi S., Petitjean P., 2001, MNRAS, 326, 597
Pichon C. et al., 2009, MNRAS, submitted
Pires S., Starck J. ., Amara A., Teyssier R., Refregier A., Fadili J., 2009, MNRAS, 395, 1265
Pogosyan D., Gay C., Pichon C., 2009a, preprint (arXiv:0907.1437)
Pogosyan D., Pichon C., Gay C., Prunet S., Cardoso J. F., Sousbie T., Colombi S., 2009b, MNRAS, 396, 635
Prunet S., Pichon C., Aubert D., Pogosyan D., Teyssier R., Gottloeber S., 2008, ApJS, 178, 179
Richardson W. H., 1972, J. Opt. Soc. America (1917–1983), 62, 55
Schirmer M., Erben T., Hetterscheidt M., Schneider P., 2007, A&A, 462
Schneider P., van Waerbeke L., Kilbinger M., Mellier Y., 2002, A&A, 396, 1
Schwarz U. J., 1978, A&A, 65, 345
Seitz S., Schneider P., Bartelmann M., 1998, A&A, 337, 325
Shapiro C., 2009, ApJ, 696, 775
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Skilling J., Strong A. W., Bennett K., 1979, MNRAS, 187, 145
Soulez F., Denis L., Thiébaut E., Fournier C., Goepfert C., 2007, J. Opt. Soc. Am. A, 24, 3708
Sousbie T., Pichon C., Courtois H., Colombi S., Novikov D., 2008a, ApJ, 672, L1
Sousbie T., Pichon C., Colombi S., Novikov D., Pogosyan D., 2008b, MNRAS, 383, 1655
Sousbie T., Colombi S., Pichon C., 2009, MNRAS, 393, 457
Starck J., Pires S., Refregier A., 2006, A&A, 451, 1139
Szapudi I., Prunet S., Colombi S., 2001, ApJ, 561, L11
Takada M., Jain B., 2002, MNRAS, 337, 875
Takada M., Jain B., 2003, MNRAS, 340, 580
Takada M., Jain B., 2004, MNRAS, 348, 897
Tarantola A., Valette B., 1982, Rev. Geophys. Space Phys., 20, 219
Teyssier R., 2002, A&A, 385, 337
Teyssier R. et al., 2009, A&A, 497, 335
Thiébaut E., 2002, in Starck J.-L., Murtagh F. D., eds, Proc. SPIE Vol. 4847, Astronomical Data Analysis II. SPIE, Bellingham, p. 174
Thiébaut E., 2005, in Foy R., Foy F. C., eds, NATO ASIB Proc. 198, Optics in Astrophysics. Springer, Amsterdam, p. 397
van Waerbeke L., Bernardeau F., Mellier Y., 1999, A&A, 342, 15
Wahba G., ed., 1990, CBMS-NSF Regional Conf. Ser. in Applied Mathematics, Spline Models for Observational Data. Society for Industrial and Applied Mathematics, Philadelphia
White M., 2005, Astropart. Phys., 23, 349

APPENDIX A: EFFICIENT STAR REMOVAL

We have observed that for realistic deep field images, GCV yields a hyper-parameter value which is relevant to regularize the higher part of the dynamic (mainly due to stars, i.e. point-like objects which concentrate their luminous energy in a very small area) but which is much too low to regularize the lower parts of the dynamic where galaxies remain. Indeed, when dealing with images with a large dynamical range, GCV yields a value of the regularization level μ which is necessarily a compromise between not smoothing too much the sharp features and sufficient smoothing of low contrasted structures to avoid noise amplification. The solution to the problem of underestimating the regularization weight can be solved by applying the GCV method on to the image with no stars. We want to find structures of known shape $s(x)$ but unknown position and intensity in the image $\boldsymbol{y}$. In our case, $s(x)$ is the PSF since we want to detect stars. This reasoning could, however, be generalized to other kind of objects. If a single object of this shape is present in the image, this could be achieved by considering the objective function:

$$\phi_{\rm full}(\alpha, t) = \sum_k w_k \left[\alpha\, s(x_k - t) - y_k\right]^2$$

to be minimized w.r.t. the weight α and the offset t, here a 2D vector. In fact, since $\boldsymbol{y}$ may be crowded with similar structures (or with other fainter structures), a better strategy is to limit the local fit to a small region of interest (ROI) around the structure. This is achieved by minimizing

$$\phi(\alpha, t) = \sum_k w_k\, r(x_k - t)\,[\alpha\, s(x_k - t) - y_k]^2\,,$$

where $r(\delta\boldsymbol{x})$ is equal to 1 within the ROI and equal to 0 outside the ROI. Minimization of $\phi(\alpha, t)$ w.r.t. α yields the best intensity for a local fit around t:

$$\frac{\partial\phi}{\partial\alpha} = 0 \quad\Longleftrightarrow\quad \alpha^\star = \frac{\sum_k w_k\, r(x_k - t)\, s(x_k - t)\, y_k}{\sum_k w_k\, r(x_k - t)\, s(x_k - t)^2}\,.$$

Inserting $\alpha^\star$ in the objective function yields

$$\phi^\star(t) \triangleq \phi(\alpha, t)|_{\alpha=\alpha^\star}\,,$$

$$= \sum_k w_k\, r(x_k - t)\, y_k^2 - \frac{\left(\sum_k w_k\, r(x_k - t)\, s(x_k - t)\, y_k\right)^2}{\sum_k w_k\, r(x_k - t)\, s(x_k - t)^2}\,.$$

Since $r(\delta\boldsymbol{x})^2 = r(\delta\boldsymbol{x})$, by defining $s_{\rm ROI}(\delta\boldsymbol{x}) \equiv r(\delta\boldsymbol{x})\, s(\delta\boldsymbol{x})$, the local criterion and local best intensity can be rewritten as

$$\phi^\star(t) = \sum_k r(x_k - t)\, w_k\, y_k^2 - \frac{\left(\sum_k s_{\rm ROI}(x_k - t)\, w_k\, y_k\right)^2}{\sum_k s_{\rm ROI}(x_k - t)^2\, w_k}\,,$$

$$\alpha^\star(t) = \frac{\sum_k s_{\rm ROI}(x_k - t)\, w_k\, y_k}{\sum_k s_{\rm ROI}(x_k - t)^2\, w_k}\,.$$

These parameters can be computed for all shifts by an integer number of pixels by means of FFTs (cross-correlation product). Unfortunately, the overall minimum of $\phi^\star(t)$ is not the best choice for removing the brightest structures since there is no warranty that this minimum corresponds to a bright object. It is better to select the location which yields the brightest structure, i.e. the maximum of $\alpha^\star(t)$. After the removal of the contribution $\alpha^\star(t^\star)\, s(x - t^\star)$ from the data, this technique can be repeated to detect the second brightest source, and so on. The corresponding algorithm is very similar to the CLEAN method (Högbom 1974; Schwarz 1978) with the further refinement of accounting for non-stationary noise and missing data. It has been shown that it achieves sub-pixel precision (Soulez et al. 2007) and that it could be used to detect (and remove) out of field sources (Soulez et al. 2007).

APPENDIX B: MODEL ON THE SPHERE

Let us describe in more details the model used for the inversion of Section 3.1.

B1 Discretization and sampling

After discretization and using explicit indices, the model in equation (17) writes

$$g_{j,k} = \frac{\gamma_{j,k}}{1 - \kappa_j} + e_{j,k}\,,$$

where the index j runs over the sky coordinates $\hat{\boldsymbol{n}}_j = (x_j, y_j)$, index k corresponds to the two components U and $\mathcal{Q}$ of the polarization, whereas ℓ and m are the harmonic indices and p refers to the two components of the spinned two-harmonic. In words, the discretization yields

$$g_{j,k} \equiv g_k(\hat{\boldsymbol{n}}_j),\quad \gamma_{j,k} \equiv \gamma_k(\hat{\boldsymbol{n}}_j),\quad \kappa \equiv \kappa(\hat{\boldsymbol{n}}_j),\quad e_{j,k} \equiv e_k(\hat{\boldsymbol{n}}_j)\,.$$

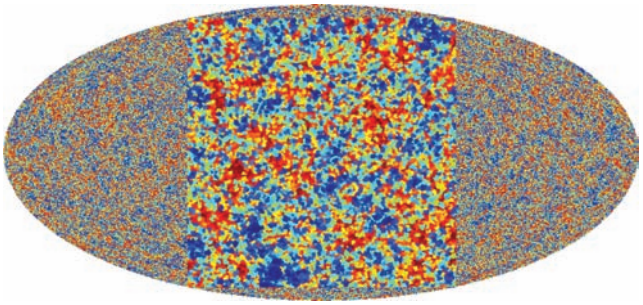

Figure B1. Peak patch of the recovered κ map. The inner box zooms the central region. The colour coding corresponds loosely to the density of the different peak patches. The PDF of the area of these patches is described in Fig. 15 while the maxima mentioned in this figure are found within each patch.

Here, the fields κ and γ are linear functions of the complex field $\boldsymbol{a}$ of the spherical harmonic coefficients of κ. Using the matrix notation of the paper, κ and γ write

$$\boldsymbol{\kappa} = \mathbf{K}\cdot\boldsymbol{a}\,,\quad \boldsymbol{\gamma} = \mathbf{G}\cdot\boldsymbol{a}\,,$$

where $\mathbf{K} = \mathbf{Y}$ and $\mathbf{G} = {}_p\mathbf{Y}\cdot\mathbf{J}$; with explicit index notations:

$$\kappa_j = \sum_{\ell,m} K_{j,\ell,m}\, a_{\ell,m} = \sum_{\ell,m} \mathbf{Y}_{j,\ell,m}\, a_{\ell,m}\,,$$

and

$$\gamma_{j,k} = \sum_{\ell,m} G_{j,k,\ell,m}\, a_{\ell,m} = \sum_{\ell,m,p} {}_p\mathbf{Y}_{j,k,\ell,m,p}\ (\mathbf{J}\cdot\boldsymbol{a})_{\ell,m,p}\,.$$

To get the detailed expression of the operator $\mathbf{J}$ we start from the relationship between the lensing potential, the convergence and the shear fields on the sphere. To do this, we need first to define the null diad, based on the polar coordinates unit vectors:

$$\mathbf{m}_\pm = \frac{(\hat{\mathbf{e}}_\theta \mp i\hat{\mathbf{e}}_\phi)}{\sqrt{2}}\,. \tag{B1}$$

Given this diad, the lensing potential, convergence and shear are related through

$$\nabla_i\nabla_j\Phi = \kappa\, g_{ij} + (\gamma_1 + i\gamma_2)(\mathbf{m}_+ \otimes \mathbf{m}_+)_{ij} + (\gamma_1 - i\gamma_2)(\mathbf{m}_- \otimes \mathbf{m}_-)_{ij}\,,$$

where g_{ij} is the spherical metric tensor, and ∇ the spherical covariant derivative. Now, using the following expression of the second covariant derivative of a scalar spherical harmonic:

$$\nabla_i\nabla_j\mathbf{Y}_{\ell,m} = \frac{1}{2}\sqrt{\frac{(l+2)!}{(l-2)!}}[{}_2\mathbf{Y}_{\ell,m}(\mathbf{m}_+ \otimes \mathbf{m}_+) + {}_{-2}\mathbf{Y}_{\ell,m}(\mathbf{m}_- \otimes \mathbf{m}_-)]_{ij} - \frac{\ell(\ell+1)}{2}\mathbf{Y}_{\ell,m}\, g_{ij}\,,$$

we can relate the convergence and shear fields to the spherical harmonic coefficients $\Phi_{\ell m}$ of the lensing potential:

$$\kappa(\hat{n}) = -\sum_{\ell m}\frac{1}{2}\ell(\ell+1)\Phi_{\ell,m}\mathbf{Y}_{\ell,m}(\hat{n})\,, \tag{B2}$$

$$(\gamma_1 \pm i\gamma_2)(\hat{n}) = \sum_{\ell m}\frac{1}{2}\sqrt{\frac{(l+2)!}{(l-2)!}}\Phi_{\ell,m\pm2}\mathbf{Y}_{\ell,m}(\hat{n})\,, \tag{B3}$$

$$= -\sum_{\ell m}(a_{\ell,m,E} \pm i a_{\ell,m,B})_{\pm2}\mathbf{Y}_{\ell,m}(\hat{n})\,, \tag{B4}$$

where the last equality defines the E and B mode coefficients. Relating the latter coefficients to the spherical harmonics decomposition of κ written above, we get the following expression for the $\mathbf{J}$ operator coefficients:

$$(\mathbf{J}\cdot\boldsymbol{a})_{\ell,m,E} = \sqrt{\frac{(\ell+2)(\ell-1)}{(\ell+1)\ell}}\, a_{\ell,m}\,, \tag{B5}$$

$$(\mathbf{J}\cdot\boldsymbol{a})_{\ell,m,B} = 0\,, \tag{B6}$$

where $a_{\ell,m}$ are the spherical harmonic coefficients of the convergence field.

B2 Likelihood

The data related term in the cost function is

$$\mathcal{L} = \sum_{j,k} w_{j,k}\left(\frac{\gamma_{j,k}}{1-\kappa_j} - g_{j,k}\right)^2 .$$

The gradient of this term is needed to find the solution of the inverse problem:

$$\frac{\partial\mathcal{L}(\boldsymbol{a})}{\partial a_{\ell,m}} = 2\sum_{j,k} H_{\ell,m,j,k}\,\frac{r_{j,k}}{1-\kappa_j} + 2\sum_j \mathbf{Y}^*_{\ell,m,j}\,\frac{\sum_k \gamma^*_{j,k} r_{j,k}}{(1-\kappa_j)^2},$$

where

$$r_{j,k} = w_{j,k}\left(\frac{\gamma_{j,k}}{1-\kappa_j} - g_{j,k}\right)$$

are the weighted residuals, and where

$$H_{\ell,m,j,k} = \sqrt{\frac{(\ell+2)(\ell-1)}{(\ell+1)\ell}}\,{}_p\mathbf{Y}^*_{\ell,m,1,j,k}\,. \tag{B7}$$

B3 Regularization

The aim of the regularization is to avoid ill-conditioning and noise amplification in the inversion. Following a Bayesian prescription, this can be achieved by requiring the field κ to obey some known a priori statistics, or while assuming a roughness penalty for $\mathcal{R}$.

B3.1 Wiener filter and ℓ_2 penalty

Assuming the field κ has Gaussian distribution with mean $\bar{\boldsymbol{\kappa}} = \langle\boldsymbol{\kappa}\rangle$ and covariance $\mathbf{C}_{\boldsymbol{\kappa}} = \langle(\boldsymbol{\kappa}-\bar{\boldsymbol{\kappa}})\cdot(\boldsymbol{\kappa}-\bar{\boldsymbol{\kappa}})^{\mathrm{T}}\rangle$, the prior penalty should write

$$\mu \equiv 1 \quad\text{and}\quad \mathcal{R} = (\boldsymbol{\kappa}-\bar{\boldsymbol{\kappa}})^{\mathrm{T}}\cdot\mathbf{C}_{\boldsymbol{\kappa}}^{-1}\cdot(\boldsymbol{\kappa}-\bar{\boldsymbol{\kappa}})\,.$$

For a field with zero mean ($\bar{\boldsymbol{\kappa}} = \mathbf{0}$) and stationary isotropic statistics, the regularization can be expressed in terms of the harmonic coefficients:

$$\mathcal{R}(\boldsymbol{a}) = \|\mathbf{C}^{-1/2}\cdot\boldsymbol{a}\|^2 = \sum_\ell \frac{\sum_m |a_{\ell,m}|^2}{C_\ell}\,, \tag{B8}$$

with

$$C_\ell = \langle|a_{\ell,m}|\rangle^2\,, \tag{B9}$$

where the angular brackets denote here the expected value taken over the index m of the harmonic coefficients. The gradient of the stationary isotropic Gaussian regularization in equation (B8) is

$$\frac{\partial\mathcal{R}(\boldsymbol{a})}{\partial a_{\ell,m}} = 2\,\frac{a_{\ell,m}}{C_\ell}\,.$$

Note that the regularization in equation (B8) with a known power spectrum C_ℓ for the field κ yields the so-called Wiener filter.

When the power spectrum of κ is not exactly known, a quadratic prior can alternatively be used. For instance,

$$\mathcal{R}(\boldsymbol{a}) = \|\mathbf{R}^{-1/2}\cdot\boldsymbol{a}\|^2 = \sum_\ell \frac{\sum_m |a_{\ell,m}|^2}{R_\ell}\,. \tag{B10}$$

In our framework, effective regularization is achieved by requiring the field κ to be somewhat smooth. In practice, this is obtained by requiring R_ℓ to be a positive non-decreasing function of the index ℓ. Note that, from a Bayesian viewpoint, the regularization in equation (B10) corresponds to the prior that κ is a stationary isotropic centred Gaussian field with mean power spectrum $C_\ell = R_\ell$, which is similar to the Wiener filter except that the exact statistics are not known in advance (because some parameters of the regularization have to be tuned; for instance, μ need not be equal to one). The gradient of $\mathcal{R}$ in equation (B10) reads

$$\frac{\partial\mathcal{R}(\boldsymbol{a})}{\partial a_{\ell,m}} = 2\,\frac{a_{\ell,m}}{R_\ell}\,.$$

The quadratic prior in equation (B10) can be expressed in terms of κ

$$\mathcal{R} = \|\mathbf{R}^{-1/2}\cdot\boldsymbol{a}\|^2 = \|\mathbf{D}\cdot\boldsymbol{\kappa}\|^2\,,$$

where $\mathbf{D} = \mathbf{R}^{-1/2}\cdot\mathbf{Y}^{\#}$ is some finite difference operator which gives an estimate of the local fluctuation of the field, and $\mathbf{Y}^{\#}$ is the (pseudo-)inverse of the scalar spherical harmonics matrix. In our framework, we choose to measure the amplitude of the local fluctuations of the field κ by its Laplacian $\nabla^2\boldsymbol{\kappa}$ and to express the regularization penalty as

$$\mathcal{R}(\boldsymbol{a}) = \sum_j \phi\left((\nabla^2\boldsymbol{\kappa})_j\right), \tag{B11}$$

where the cost function $\phi(r)$ is an increasing function of $|r|$. When $\phi(r) = r^2$, our regularization is a quadratic penalty similar to equation (B10). Using matrix notation, the Laplacian of the field $\boldsymbol{\kappa}$ writes

$$\nabla^2\boldsymbol{\kappa} = \mathbf{Y}\cdot\mathbf{L}^{-1/2}\cdot\boldsymbol{a}\,,\quad\text{with}\quad \left(\mathbf{L}^{-1/2}\cdot\boldsymbol{a}\right)_{\ell,m} = \frac{a_{\ell,m}}{\sqrt{L_\ell}}\,,$$

where $L_\ell \equiv \ell^{-2}(\ell+1)^{-2}$. In order to perform the minimization, the gradient of the regularization must be computed. By the chain rule

$$\begin{aligned}\frac{\partial\mathcal{R}(\boldsymbol{a})}{\partial a_{\ell,m}} &= \sum_j \phi'\left((\nabla^2\boldsymbol{\kappa})_j\right)\,\frac{\partial(\nabla^2\boldsymbol{\kappa})_j}{\partial a_{\ell,m}}\,,\\ &= \sum_j \frac{\mathbf{Y}^*_{\ell,m,j}}{\sqrt{L_\ell}}\,\phi'\left((\nabla^2\boldsymbol{\kappa})_j\right),\end{aligned} \tag{B12}$$

where $\phi'(r)$ is the derivative of $\phi(r)$.

B3.2 $\ell_2 - \ell_1$ penalty

As for the image restoration, quadratic regularization yields spurious ripples in the regularized κ map. To avoid them, we propose to use a $\ell_2 - \ell_1$ cost function ϕ applied to the Laplacian of κ. The details of the $\ell_2 - \ell_1$ cost function are discussed in Section 2.1.4. Taking $\mathcal{R}(\boldsymbol{a}) = \sum_j \phi\left((\nabla^2\boldsymbol{\kappa})_j\right)$, with ϕ given in equation (16), yields

$$\begin{aligned}\frac{\partial\mathcal{R}(\boldsymbol{a})}{\partial a_{\ell,m}} &= \sum_j \frac{2\,\varepsilon\,(\nabla^2\boldsymbol{\kappa})_j}{\varepsilon + \left|(\nabla^2\boldsymbol{\kappa})_j\right|}\,\frac{\partial(\nabla^2\boldsymbol{\kappa})_j}{\partial a_{\ell,m}}\,,\\ &= 2\,\varepsilon\sum_j \frac{\mathbf{Y}^*_{\ell,m,j}}{\sqrt{C_\ell}}\,\frac{(\nabla^2\boldsymbol{\kappa})_j}{\varepsilon + \left|(\nabla^2\boldsymbol{\kappa})_j\right|}\,.\end{aligned}$$

In practice, we use GCV to set the level of the regularization, possibly after cluster removal (as explained in Appendix A) and the $\ell_1 - \ell_2$ threshold is set to be $\varepsilon = \alpha\,\sigma$ where $\alpha \sim 2$–3 and σ is the standard deviation of the histogram of spatial finite differences.

APPENDIX C: FROM THE SPHERE TO THE PLANE

In Section 3.1.2, we sketched the correspondence between the full-sky and the flat-sky approximation of the lens equation. Let us derive it here precisely and use it to investigate the effect of shot noise in the estimation of κ.

C1 Derivation

Following closely Hu (2000), let us start with a scalar field on the sphere, and its decomposition on the usual spherical harmonics:

$$X(\hat{n}) = \sum_{\ell m} X_{\ell,m} \mathbf{Y}_{\ell,m}\,, \tag{C1}$$

and let us define

$$X(\boldsymbol{l}) = \sqrt{\frac{4\pi}{2\ell+1}} \sum_m i^{-m} X_{\ell,m} \mathrm{e}^{\mathrm{i}m\phi_\ell}\,, \tag{C2}$$

together with the inverse relation

$$X_{\ell,m} = \sqrt{\frac{2\ell+1}{4\pi}} i^m \int \frac{\mathrm{d}\phi_\ell}{2\pi} X(\boldsymbol{l}) \mathrm{e}^{-\mathrm{i}m\phi_l}\,,$$

where ϕ_ℓ is the polar angle of the $\boldsymbol{l}$ vector in Fourier space. Let us show that $X(\boldsymbol{l})$ corresponds to the Fourier decomposition of the field in the flat-sky limit (small angles near the pole). Indeed, taking the asymptotic behaviour of the spherical harmonics

$$\mathbf{Y}_{\ell,m} \approx J_m(\ell\theta)\sqrt{\frac{\ell}{2\pi}} \mathrm{e}^{\mathrm{i}m\phi}\,,$$

together with the plane-wave expansion in terms of Bessel functions

$$e^{i\boldsymbol{l}.\hat{n}} = \sum_m i^m J_m(\ell\theta) \mathrm{e}^{\mathrm{i}m(\phi-\phi_\ell)} \approx \sqrt{\frac{2\pi}{\ell}} \sum_m i^m \mathbf{Y}_{\ell,m} \mathrm{e}^{\mathrm{i}m\phi_l}\,.$$

We get from equation (C1)

$$\begin{aligned} X(\hat{n}) &\approx \sum_\ell \frac{\ell}{2\pi} \int \frac{\mathrm{d}\phi_\ell}{2\pi} X(\boldsymbol{l}) \sum_m J_m(\ell\theta) i^m \mathrm{e}^{\mathrm{i}m(\phi-\phi_\ell)}\,, \\ &\approx \int \frac{\mathrm{d}^2\ell}{(2\pi)^2} X(\boldsymbol{l}) \mathrm{e}^{\mathrm{i}\boldsymbol{l}.\hat{n}}\,. \end{aligned}$$

For a spin-2 field, let us proceed similarly. We start from the all-sky definition of a spin-2 tensor field, and its decomposition in spin-2 spherical harmonics:

$${}_\pm X(\hat{n}) = \sum_{\ell m} {}_\pm X_{\ell,m} {}_{\pm 2}\mathbf{Y}_{\ell,m}\,, \tag{C3}$$

where ${}_\pm X(\hat{n})$ is defined in the spherical tangent coordinates e_θ, e_ϕ. We define, as in equation (C2), the Fourier modes of the components of the spin-2 field as ${}_\pm X(\boldsymbol{l})$. We have in the flat-sky limit the following asymptotic form for the spin-2 spherical harmonics:

$${}_{\pm 2}\mathbf{Y}_{\ell,m} \approx \frac{1}{\ell^2} e^{\mp 2i\phi} (\partial_x \pm i\partial_y)^2 \mathbf{Y}_{\ell,m}\,. \tag{C4}$$

Plugging equation (C3) into equation (C4) yields

$$\begin{aligned} {}_\pm X(\hat{n}) &\approx \sum_\ell \frac{\ell}{2\pi} \int \frac{\mathrm{d}\phi_\ell}{2\pi} X(\boldsymbol{l}) \mathrm{e}^{\mp 2\mathrm{i}\phi} \frac{1}{\ell^2} (\partial_x \pm i\partial_y)^2 \mathrm{e}^{\mathrm{i}\boldsymbol{l}.\hat{n}}\,, \\ &\approx -\int \frac{\mathrm{d}^2\ell}{(2\pi)^2} {}_\pm X(\boldsymbol{l}) \mathrm{e}^{\pm 2\mathrm{i}(\phi_\ell-\phi)} \mathrm{e}^{\mathrm{i}\boldsymbol{l}.\hat{n}}\,. \end{aligned}$$

Redefining the spin-2 field in the fixed coordinate system such that the first axis (e_x) is aligned with $\phi = 0$, we obtain

$${}_\pm X'(\hat{n}) \approx -\int \frac{\mathrm{d}^2\ell}{(2\pi)^2} {}_\pm X(\boldsymbol{l}) \mathrm{e}^{\pm 2\mathrm{i}\phi_\ell} \mathrm{e}^{\mathrm{i}\boldsymbol{l}.\hat{n}}\,, \tag{C5}$$

where $\ell_x + i\ell_y = \ell e^{i\phi_\ell}$. Expanding ${}_\pm X(\boldsymbol{l}) = E(\boldsymbol{l}) \pm iB(\boldsymbol{l})$, we can relate these rotationally invariant quantities to the Fourier transforms of the spin-2 field individual components. In the case of weak lensing, we get the following flat-sky limits:

$$\kappa(\hat{n}) \approx -\frac{1}{2} \int \frac{\mathrm{d}^2\ell}{(2\pi)^2} \ell^2 \Phi(\boldsymbol{l}) \mathrm{e}^{\mathrm{i}\boldsymbol{l}.\hat{n}} \tag{C6}$$

$$(\gamma_1 \pm i\gamma_2)'(\hat{n}) \approx -\frac{1}{2} \int \frac{\mathrm{d}^2\ell}{(2\pi)^2} \ell^2 \Phi(\boldsymbol{l}) \mathrm{e}^{\pm 2\mathrm{i}\phi_\ell} \mathrm{e}^{\mathrm{i}\boldsymbol{l}.\hat{n}}\,. \tag{C7}$$

After identification, we thus get the limits for the operator $\mathbf{J}$:

$$\mathbf{J} = (\mathbf{1}, \mathbf{0}) \tag{C8}$$

independently of the Fourier mode modulus.

C2 S/N investigation in the plane

Let us briefly investigate the effect of noise on the recovery of the κ power spectrum arising from the finite number of sources per unit area. For this purpose, let us consider the simplest setting corresponding to a Cartesian map without mask which can therefore be inverted linearly following equations (22)–(23). In this regime, the regularized solution is simply given in Fourier space by

$$\hat{\kappa} = \frac{1}{1+\mu\left(\ell_x^2+\ell_y^2\right)} \left[\hat{g}_x \frac{\left(\ell_x^2-\ell_y^2\right)}{\left(\ell_x^2+\ell_y^2\right)} + \frac{2\ell_x\ell_y}{\left(\ell_x^2+\ell_y^2\right)} \hat{g}_y \right]\,, \tag{C9}$$

where $\hat{g}_x$, $\hat{g}_y$ and $\hat{\kappa}$ are the Fourier transform of the observed shear and convergence, and μ the penalty hyper-parameter. In Fig. C1, we make use of the simulation ${}_{1024}\,C_{\rm lin}^\ell$, whose residues (after non-linear

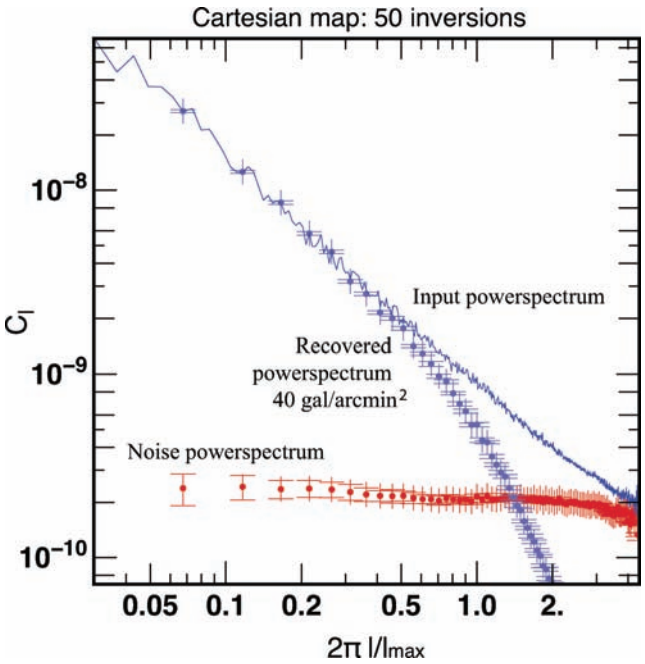

Figure C1. The effect of noise on the reconstruction of the power spectrum for a set of 50 realizations of the noise for the map ${}_{1024}\,C_{\rm lin}^\ell$ ($\ell_{\rm max} = 1200$). Note that at these scales, the spread in the recovered power spectra for the different realizations is only visible above the cut-off frequency.

inversion) are shown in Fig. 8. Here, 50 Monte Carlo realizations of the noise corresponding 40 galaxies/□arcmin are averaged to produce an estimate of the corresponding errors. Clearly, the shot noise remains small at all considered frequencies.

APPENDIX D: CONVERGENCE MAPS

The inversion technique described in the main text was validated using the mocks extracted from the HORIZON-4π simulation (Prunet et al. 2008). Let us briefly describe here how this simulation was used to generate mock slices and κ maps.

D1 Light cone generation

The generation of a light cone during run time can be performed easily at each coarse time-step of the simulation. Given a choice of the observer position in the simulation box, that we suppose here for simplicity to be at the origin of coordinates, it is easy to select the particles that belong to the slice in between redshifts $z_2 < z_1$ corresponding to two successive coarse time-steps: if (x, y, z) are the comoving coordinates of a particle, and $d = \sqrt{x^2 + y^2 + z^2}$ its comoving distance from the observer, we must have $d_{\rm dist}(z_2) < d \leq d_{\rm dist}(z_1)$ for the particle to be selected, where $d_{\rm dist}(z)$ is the comoving distance that a photon covers between redshift z and present time in the simulation box: $d_{\rm dist} = \int c{\rm d}t/a(t)$, where c is the speed of light and a the expansion factor. The problem is that structures evolve during a coarse time-step, so there are necessarily some discontinuities at the border between two successive light cone slices. These discontinuities are due to large-scale motions of particles plus their thermal velocity within dark matter haloes. Given the large size of the simulation considered here, thermal motion within the largest cluster is expected to bring the most significant effects of discontinuity. For a particle with peculiar velocity v, the largest discontinuity to be expected, i.e. the largest possible difference between expected and actual position of the particle, is given by

$$\Delta = (v/c)[d_{\rm dist}(z_1) - d_{\rm dist}(z_2)]. \tag{D1}$$

In equation (D1), we performed a linear Lagrangian approximation, i.e. we neglected variations of the velocity of the particle during the coarse time-step. Using the Press & Schechter formalism, or the improved formula of Sheth & Tormen (1999), the mass of the largest cluster in the HORIZON simulation solves approximately the implicit equation

$$\Omega_0 \rho_{\rm c} L^3 F[M_{\rm max}(z), z]/M_{\rm max}(z) = 1, \tag{D2}$$

where $\rho_{\rm c}$ is the critical density of the Universe and F is the fraction of mass in the Universe in objects of mass larger than M. Basically, this equation states that the mass in objects of mass larger than M is equal to M, which means that we are left with only one cluster of mass M, the largest detectable cluster in our cube of size L. We can compute $F(M, z)$ with the usual formula, e.g.

$$F(M, z) = \int_{\mu > \nu(M,z)} f(\mu){\rm d}\mu, \tag{D3}$$

with $\nu = 1.686/\sigma(M, z)$ where $\sigma(M, z)$ is the linear variance at redshift z corresponding to mass scale M, and $f(\mu)$ is given by equation (10) of Sheth & Tormen (1999). Performing these calculations, we find that the largest cluster at present time in a cube of size $L = 2000\,h^{-1}$ Mpc should have a typical mass of $M_{\rm max}(z = 0) \simeq 1.47 \times 10^{15}\,{\rm M}_\odot$. With a standard Friend-of-Friend algorithm using a linking parameter $b = 0.2$, we find that the most massive halo detected in the simulation presents a somewhat larger mass, $M = 5.4 \times 10^{15}\,{\rm M}_\odot$. Yet, in that rare events regime, we cannot expect our theoretical estimate to be more accurate. What matters, though, is the thermal velocity rather than the mass. Applying the Virial theorem, we have (e.g. Peacock 1999) $v^2 \simeq GM_{\rm max}/R_{\rm vir}$, with

$$\frac{4}{3}\pi R_{\rm vir}^3 \rho_{\rm vir} = M_{\rm max}, \quad \rho_{\rm vir} \simeq 178\Omega_0\rho_{\rm c}(1+z)^3/\Omega(z)^{0.7},$$

where $\Omega(z)$ is the density parameter as a function of redshift ($\Omega(0) \equiv \Omega_0$). These expressions are given in physical coordinates hence the factor $(1 + z)^3$ in the expression of $\rho_{\rm vir}$. This reads, at $z = 0$, $v \simeq 1570$ km s^{-1} for $M_{\rm max}(z = 0) \simeq 1.47 \times 10^{15}\,{\rm M}_\odot$. In the largest cluster of the simulation, the overall velocity dispersion is of the order of 2100 km s^{-1}, a slightly larger value that reflects the actual value of the mass. To be conservative, we estimate the expected errors in equation (D1) with the Virial velocity rescaled by a factor of 2100/1570, and with a further multiplication by a factor of 3 to be in the 3σ regime. The corresponding maximal expected discontinuity displacement is shown in Mpc as a function of the expansion factor in Fig. D1. As expected from the dynamically self-consistent calculation of the coarse time-step (which is basically determined by a Courant condition using the velocity field), the comoving error does not change significantly with redshift and remains below the very conservative limit of 200 kpc. Obviously, we expect in practice the errors brought by discontinuities to be in general much smaller than that, as for $z = 0$ the present errors correspond to unrealistic velocities as large as about 6000 km s^{-1}!

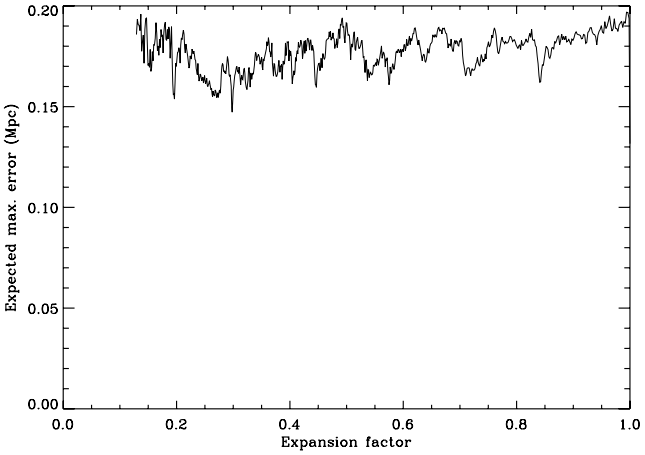

Figure D1. The expected maximum uncertainty on particle positions due to the method used to create the light cone as a function of the expansion factor. It is computed according to equation (D1) with a velocity v estimated to be three times the Virial velocity of the largest cluster in the simulation.

D2 From slices to κ maps

In the main text, the expression for κ as a function of the density contrast in the simulation is given in equation (28) in the geometric optic approximation. Let us rearrange this formula in a form that is more suited to integration over redshift slices in a simulation.

$$\kappa(\hat{\mathbf{n}}_{\rm pix}) \approx \frac{3}{2}\Omega_{\rm m} \sum_b W_b \frac{H_0}{c} \int_{\Delta z_b} \frac{c{\rm d}z}{H_0 E(z)} \delta\left(\frac{c}{H_0}\mathcal{D}(z)\hat{\mathbf{n}}_{\rm pix}, z\right),$$

where

$$W_b = \left(\int_{\Delta z_b} \frac{{\rm d}z}{E(z)} \frac{\mathcal{D}(z)\mathcal{D}(z, z_{\rm s})}{\mathcal{D}(z_{\rm s})} \frac{1}{a(z)}\right) \Bigg/ \left(\int_{\Delta z_b} \frac{{\rm d}z}{E(z)}\right)$$

is a slice-related weight, and the integral over the density contrast, δ, reads

$$I = \int_{\Delta z_b} \frac{c\mathrm{d}z}{H_0 E(z)} \delta\left(\frac{c}{H_0}\mathcal{D}(z)\hat{\mathbf{n}}_{\rm pix}, z\right),$$
$$= \int_{\Delta \chi_b} \mathrm{d}\chi\, \delta(\chi \hat{n}_{\rm pix}, \chi) \approx \frac{V(\mathrm{simu})}{N_{\rm part}(\mathrm{simu})}\left(\frac{N_{\rm part}(\theta_{\rm pix}, z_b)}{S_{\rm pix}(z_b)} - 1\right),$$

where

$$S_{\rm pix}(z_b) = \frac{4\pi}{N_{\rm pix}}\frac{c^2}{H_0^2}\mathcal{D}^2(z_b)$$

is the comoving surface of the spherical pixel. Putting all together, we get the following formula for the convergence map:

$$\kappa(\theta_{\rm pix}) = \frac{3}{2}\Omega_{\rm m}\frac{N_{\rm pix}}{4\pi}\left(\frac{H_0}{c}\right)^3 \frac{V(\mathrm{simu})}{N_{\rm part}(\mathrm{simu})}\sum_b W_b \frac{N_{\rm part}(\theta_{\rm pix}, z_b)}{\mathcal{D}^2(z_b)}.$$

Once the κ map is available, it is straightforward to build the corresponding **ssg** using equation (17).

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