Probability distribution function of the aperture mass field with large deviation theory

Alexandre Barthelemy\textsuperscript{1,*}, Sandrine Codis\textsuperscript{1}, Francis Bernardeau\textsuperscript{1,2}

\textsuperscript{1}CNRS & Sorbonne Université, UMR 7095, Institut d’Astrophysique de Paris, 75014, Paris, France
\textsuperscript{2}Institut de Physique Théorique, Université Paris-Saclay, CEA, CNRS, UMR 3681, 91191 Gif-sur-Yvette, France

ABSTRACT
In the context of tomographic cosmic shear surveys, a theoretical model for the one-point statistics of the aperture mass (Map) is developed. This formalism is based on the application of the large deviation principle to the projected matter density field and more specifically to the angular aperture masses. The latter holds the advantage of being an observable that can be directly extracted from the observed shear field and to be, by construction, independent from the long wave modes. Furthermore we show that, with the help of a nulling procedure based on the so-called BNT transform, it is possible to build observables that depend only on a finite range of redshifts making them also independent from the small-scale modes. This procedure makes predictions for the shape of the one-point Probability Distribution Function of such an observable very accurate, comparable to what had been previously obtained for 3D observables. Comparisons with specific simulations reveal however inconsistent results showing that synthetic lensing maps were not accurate enough for such refined observables. It points to the need for more precise dedicated numerical developments whose performances could be benchmarked with such observables. We furthermore review the possible systematics that could affect such a formalism in future weak-lensing surveys like Euclid, notably the impact of shape noise as well as leading corrections coming from lens-lens couplings, geodesic deviation, reduced shear and magnification bias.

Key words: cosmology: theory – large-scale structure of Universe – gravitational lensing: weak – methods: analytical, numerical

1 INTRODUCTION
The effect of weak gravitational lensing (WL) originates from the propagation of light rays through the inhomogeneous distribution of baryonic and dark matter which induces slight (de)magnification of the brightness of galaxies and distortion from their intrinsic shape (Kaiser 1992). The statistics of WL fields provide a powerful tool for precision cosmology (see e.g. a review in Kilbinger (2015)) and motivated the build up of new generation large galaxy surveys such as the Legacy Survey of Space and Time (LSST) (Ivezić et al. 2019) or Euclid (Laureijs et al. 2011) which will provide data of unprecedented quality in the coming years. As such and in order to reach the percent precision and accuracy on the estimation of cosmological parameters, theorists need to build tools that can optimally extract information from those datasets and be able to provide accurate predictions in the non-linear regime of cosmic structure formation.

The most common approach is to focus on the information contained in the power spectra or equivalently their real-space counterparts the two-point correlation functions. Unfortunately these observables contain only complete statistical information for Gaussian random fields, a prescription valid with extremely good accuracy to describe primordial metric perturbations visible in the cosmic microwave background (Planck Collaboration et al. 2020). However, even starting from Gaussian initial conditions, the subsequent non-linear time-evolution of density fluctuations by means of the gravitational instability develops significant non-Gaussianities, in particular for small scales and late times. In this non-linear regime of structure formation, we observe both an increase in power in the power spectrum measurements (relative to linear evolution) and a generation of distinct non-Gaussianities in the late-time density field, which by projection also implies strong non-Gaussian features of the weak-lensing fields carrying non-negligible cosmological information. More quantitatively, Zorrilla Matilla et al. (2020) recently showed with a series of Deep Neural Networks how a significant amount of the cosmological information in the convergence field lies in the extreme rare events, that is the tails of the probability distribution function (PDF), and a Fisher analysis based on fast simulations in Patton et al. (2017) demonstrated that the weak-lensing convergence PDF provides information complementary to the cosmic shear two-point correlation. Still making use of Fisher analysis but this time using a “first principles” theoretical model of the con-
vergence PDF, Boyle et al. (2020) also found that it provides tighter constraints for the equation of state of dark energy, the amplitude of fluctuations, the total matter fraction and the sum of neutrino masses, especially when performing a multi-scale analysis and in addition to the two-point correlation function. For the case of other non-Gaussian statistics, namely peak counts and Minkowski functionals, a recent analysis can be found in Zürcher et al. (2020) where it is shown that the Figure-of-Merit in the $\Omega_m-\sigma_8$ plane increases by a factor of 5 when adding those non-Gaussian statistics to the standard angular power spectrum. Note however that these quantitative analysis assume that non-Gaussian quantities can be predicted with great precision at the chosen filtering scales, which is not all given especially when using un-tested – because of the absence of a good theoretical model – predictions from numerical simulations.

Because convergence maps are always only reconstructed up to a mass sheet degeneracy, Schneider et al. (1998) introduced the aperture mass field which boils down to filtering the convergence field with a compensated filter which also possesses a dual representation in the (almost) directly observable tangential-shear space. The study of the aperture mass can thus be performed directly from the measured data and treatments of masks in the field of view of a specific survey can for example be more controlled (Porth et al. 2020; Bard et al. 2016). Moreover the same non-Gaussian statistics (peaks, moments, PDF to name a few) can be used to probe the non-Gaussian features of the aperture mass field and are found, similarly to the convergence field, to provide complementary information to the power spectrum in different surveys (Martinet et al. 2018; Kacprzak et al. 2016) and very recently in Martinet et al. (2020).

In this paper we build a theoretical model for the aperture mass PDF. So far, only a few theoretical developments have been carried out in this direction: in the early 2000’s Bernardou & Valageas (2000) and following papers or Munshi et al. (2004) built the $M_{ap}$ PDF assuming different hierarchic models for the underlying density field. More recently Reimberg & Bernardou (2018) used, as in this paper, large deviation theory (Bernardeau & Reimberg 2016) to compute the reduced-shear correction to the $M_{ap}$ PDF, but without accounting for the geometry of the past light-cone and projection effects nor comparing the predictions to numerical simulations, which are the main purposes of the present work.

The paper is organised as follows. Section 2 introduces all the formalism and procedure necessary to compute the aperture mass PDF with large deviation theory and the nulling procedure we use to reduce the sensitivity to very non-linear scales and baryonic effects. Section 3 describes the numerical simulations we use for comparison with our theoretical model and discusses diverse challenges we encountered. Section 4 performs a succinct evaluation of the impact of shape noise, assuming a Euclid-like instrument, and cosmic variance on the aperture mass PDF thus giving an idea of the level of accuracy that theoretical models need to reach. Finally, section 5 concludes. We give in the successive appendices many technical details on the procedure applied here, discuss the extension of the formalism to other filters than top-hats and also estimate the leading-order corrections coming from couplings between lenses, geodesic deviation, reduced shear and the magnification bias.

## 2 APERTURE MASS PDF

### 2.1 Aperture mass definition

The convergence $\kappa$ can be interpreted as a line-of-sight projection of the matter density distribution between the observer and the source. More quantitatively it can be written as (Mellier 1999)

$$\kappa(\vartheta) = \int_0^{\chi_s} d\chi \, \omega(\chi, \chi_s) \delta(\chi, D\vartheta),$$

(1)

where $\chi$ is the comoving radial distance – $\chi_s$ the radial distance of the source – that depends on the cosmological model, and $D$ is the comoving angular distance

$$D(\chi) \equiv \left\{ \begin{array}{ll}
\frac{\sin(\sqrt{K}\chi)}{\sqrt{K}} & \text{for } K > 0 \\
\chi & \text{for } K = 0 \\
\frac{\sinh(\sqrt{-K}\chi)}{\sqrt{-K}} & \text{for } K < 0
\end{array} \right.,$$

(2)

with $K$ the constant space curvature. The lensing kernel $\omega$ is defined as

$$\omega(\chi, \chi_s) = \frac{3 \Omega_m H_0^2}{2 c^2} \frac{D(\chi) D(\chi_s - \chi)}{D(\chi_s)} (1 + z(\chi)).$$

(3)

Note that equation (1) assumed no lens-lens couplings and no geodesic perturbations (Born approximation). Although they could affect the higher-order (joint) cumulants we are computing and the convergence PDF, it was showed analytically in Bernardou et al. (1997) and numerically in Petri et al. (2017) that the effect on the convergence skewness for the sources and scales of interest is negligible. A prescription to include those effects, mostly relevant in the context of CMB lensing, in the depicted formalism i.e large deviation theory was given in Barthelemy et al. (2020b).

The aperture mass $M_{ap}$ is defined as a geometrical average of the local convergence with a window of vanishing average

$$M_{ap}(\vartheta) = \int d^2 \vartheta’ U_\vartheta(\vartheta’) \kappa(\vartheta’ - \vartheta)$$

(4)

with

$$\int d^2 \vartheta’ U_\vartheta(\vartheta’) = 0.$$  

(5)

Because convergence maps are always only reconstructed up to a mass sheet degeneracy, statistical quantities that can be measured in terms of convergence maps and which are not affected by this degeneracy can only be smoothed quantities with compensated filters as is the case for aperture mass maps. Moreover, aperture mass can be interestingly expressed as a function of the tangential component $\gamma_t$ of the shear (Kaiser 1995; Schneider 1996)

$$M_{ap}(\vartheta) = \int d^2 \vartheta’ Q_\vartheta (\vartheta’) \gamma_t (\vartheta - \vartheta’),$$

(6)

where

$$Q_\vartheta(\vartheta) = -U_\vartheta(\vartheta) + \frac{2}{\vartheta^2} \int_0^\vartheta d\vartheta’ \vartheta’ U_\vartheta (\vartheta’),$$

(7)

thus rendering the aperture mass a direct observable up to a reduced shear correction but which can be accounted for as discussed in appendix F.
In principle, the large deviation formalism could be used with any filter function as was shown in Bernardes & Reimberg (2016), Reimberg & Bernardes (2018) and in appendix E. However, we will here adopt a simple prescription of the 2D density slope obtained via large-deviation theory. Using this result, we will then build the non-linear cumulant generating function of the aperture mass and its PDF. A schematic representation of our procedure can be found in Fig. 2.

2.2 Projection formula

Under small-angle/Limber approximation, it has been showed in Bernardes & Valageas (2000); Barthelemy et al. (2020a) that the convergence field filtered in an angular top-hat window function can be seen as a juxtaposition of statistically independent 2D slices of the underlying density field. Filtering the field using equation (8) does not change the demonstration and thus the cumulants of the aperture mass lensing field are given by

$$ M_{\text{ap}}(\theta) = \kappa_{<\theta_2}(\theta) - \kappa_{<\theta_1}(\theta), $$

where $\kappa_{<\theta_{1,2}}$ denotes the convergence field filtered by a top-hat window of angular radius $\theta_1$ and $\theta_2 = 2\theta_1$. This choice is both motivated by the relative simplicity of obtaining statistics of concentric disks/spheres within the large deviation formalism – other choices of compensated filters are often used in the literature (see for starter Schneider et al. (1998)) – and also because top-hat filtering allows for a more rigorous assessment of the scales correctly described by the theory presented in this paper. We show the shape of our filter in Fig. 1 compared to the one used in Schneider et al. (1998).

2.3 Generating functions and PDF

Throughout this work, we make use of different statistical quantities that we briefly introduce here for clarity. From the PDF $P_X$ of some continuous random variable $X$ one can define the moment generating function as the Laplace transform of the PDF

$$ M_X(\lambda) = \mathbb{E}(e^{\lambda X}) = \int_{-\infty}^{+\infty} e^{\lambda x} P_X(x) dx, $$

or equivalently as the expectation value\(^1\) of the random variable $e^{\lambda X}$. The moment generating function, as its name implies, can be used to find the moments of the distribution as can be seen from the series expansion of the expectation of

\(^1\) Note that we make use throughout this work of the ergodicity hypothesis where one assumes that ensemble averages are equivalent to spatial averages ($\mathbb{E}() \rightarrow \langle \rangle$) over one realisation of a random field at one fixed time. This requires that spatial correlations decay sufficiently rapidly with separation such that one has access to many statistically independent volumes in one realisation.
\[ e^{\lambda X}, \]

\[ M_X(\lambda) = \mathbb{E}(e^{\lambda X}) = 1 + \lambda \mathbb{E}(X) + \frac{\lambda^2 \mathbb{E}(X^2)}{2!} + \frac{\lambda^3 \mathbb{E}(X^3)}{3!} + \cdots \]

\[ = \sum_{n=0}^{+\infty} \frac{\lambda^n \mathbb{E}(X^n)}{n!}, \tag{11} \]

so that the \( n \)-th derivative of the moment generating function in \( \lambda = 0 \) is equal to the \( n \)-th order moment, \( \mathbb{E}(X^n) \). The logarithm of the moment generating function is the cumulant generating function (CGF)

\[ \phi_X(\lambda) = \log(M_X(\lambda)) = \sum_{n=1}^{+\infty} \frac{k_n}{n!} \lambda^n \]

where \( k_n \) are the cumulants (i.e the connected moments) of the distribution.

These definitions can of course be extended to the case of joint probabilities. For the case of two continuous random variables \( X \) and \( Y \) relevant to this paper, the joint cumulant generating function reads

\[ \phi_{X,Y}(\lambda_1, \lambda_2) = \log(M_{X,Y}(\lambda_1, \lambda_2)) = \log \left[ \mathbb{E}(e^{\lambda_1 X + \lambda_2 Y}) \right] \]

\[ = \sum_{p,q=0}^{+\infty} \langle X^p Y^q \rangle \epsilon \frac{\lambda_1^p \lambda_2^q}{p! q!} - 1, \tag{13} \]

which in particular allows us to straightforwardly define the CGF of any linear combination of random variables from their joint CGF. More precisely, noticing that \( \phi_{X,Y}(\lambda) = \phi_X(\lambda, \lambda) \) allows us to generalise the famous relation

\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y) \]

to any cumulant

\[ k_{n,X+Y} = \sum_{j=0}^{n} \binom{n}{j} \langle X^j Y^{n-j} \rangle \epsilon. \tag{14} \]

It turns out that the quantities

\[ S_n = \frac{k_n}{k_2^{n-1}}, \tag{15} \]

called reduced cumulants and where \( k_2 \) is the variance, are of importance in our context as the ratios \( S_n \) of the cosmic matter density field were indeed shown to be independent from the variance (and therefore redshift) down to mildly non-linear scales (Peebles 1980; Baugh et al. 1995). We thus also define the scaled cumulant generating function (SCGF hereafter) as

\[ \varphi(\lambda) = \lim_{k_2 \to 0} \sum_{n=0}^{+\infty} \frac{S_n}{n!} \lambda^n = \lim_{k_2 \to 0} k_2 \phi_X \left( \frac{\lambda}{k_2} \right), \tag{16} \]

that we will in our context extrapolate to non-zero values of the variance. Eventually, from the CGF, one can then reconstruct the PDF for the random variable \( X \) as an inverse Laplace transform (inverting equation (10)) given by

\[ \mathcal{P}_X(x) = \int_{-\infty}^{+\infty} \frac{d\lambda}{2\pi i} \exp(-\lambda x + \phi_X(\lambda)). \tag{17} \]

### 2.4 Statistics of the 2D density slope obtained via large deviation theory

Let us now recall some of the results of LDT for matter densities in disks (equivalently long cylinders). For more details, we refer the reader to Bernardè & Reimberg (2016) and Uhlemann et al. (2018b). We will first give general notions on random variables admitting a large deviation principle then move on to the specific case of the matter density field.

A set of random variables \( \{\rho_i\} \), \( 1 \leq i \leq N \), with joint PDF \( \mathcal{P}_{\epsilon}(\{\rho_i\}) \) is said to satisfy a large deviation principle if the limit

\[ \Psi(\{\rho_i\}) = -\lim_{\epsilon \to 0} \epsilon \log[\mathcal{P}_{\epsilon}(\{\rho_i\})] \tag{18} \]

exists, where \( \epsilon \) is the driving parameter. This driving parameter indexes the set of random variables with respect to some evolution, for example a time evolution. For example, a common example of a random variable satisfying a large deviation principle is the sum of successive coin tosses where the driving parameter is one over the number of tosses entering the sum. In the case of the matter density field at a single scale this driving parameter is its variance which acts as a clock from initial to late times. For the joint statistics of concentric disks of matter, the common driving parameter could be the variance at any radius/scale since all variances behave the same in the 0 limit, being proportional to the growth rate of structure in the linear regime such that the SCGF – or equivalently cumulants at tree order – is not affected by this choice. We now omit the \( \epsilon \) sub/superscripts in our notation for simplicity.

The existence of a large deviation principle for the set of random variables \( \{\rho_i\} \) implies that their SCGF \( \varphi(\rho_i) \) is given through Varadhan’s theorem as the Legendre-Fenchel transform of the rate function \( \Psi(\rho_i) \)

\[ \varphi(\{\rho_i\}) = \varphi(\{\lambda_i\}) = \sup_{\{\lambda_i\}} \left\{ \sum \lambda_i \rho_i - \Psi(\rho_i)\{\lambda_i\} \right\}, \tag{19} \]

where the Legendre-Fenchel transform reduces to a simple Legendre transform when \( \Psi(\rho_i) \) is convex. In that case,

\[ \varphi(\{\rho_i\}) = \sum \lambda_i \rho_i - \Psi(\rho_i)\{\rho_i\}, \tag{20} \]

where \( \{\rho_i\} \) are a function of \( \{\lambda_i\} \) through the stationary conditions

\[ \lambda_k = \frac{\partial \Psi(\rho_i)\{\rho_i\}}{\partial \rho_k}, \quad \forall k \in \{1, \cdots, N\}. \tag{21} \]

Another consequence of the large-deviation principle is the so-called contraction principle. This principle states that for a set of random variables \( \{\tau_i\} \) satisfying a large deviation principle and related to \( \{\rho_i\} \) through the continuous map \( f \), then the rate function of \( \{\rho_i\} \) can be computed as

\[ \Psi(\rho_i)\{\rho_i\} = \inf_{\{\tau_i\}:f(\{\tau_i\})=\{\rho_i\}} \Psi(\tau_i)\{\tau_i\}. \tag{22} \]

This formula is called the contraction principle because \( f \) can be many-to-one in which case we are contracting information about the rate function of one random variable down to the other. In physical terms, this states that an improbable fluctuation of \( \{\rho_i\} \) is brought about by the most probable of all improbable fluctuations of \( \{\tau_i\} \).
For the case of the matter density field and starting from Gaussian initial conditions\(^2\), the rate function of the linear field is simply given by a quadratic term. Using the contraction principle, the rate function of the late-time density field at different scales can then be computed from the initial conditions if the most likely mapping between the two is known, that is if one is able to identify the leading field configuration that will contribute to the infimum of equation (22). In cylindrically symmetric configurations, which is the case for a disk of radius \(R_L\) in a slice at redshift \(z\), one could conjecture (Valageas 2002) that the most likely mapping between initial and final conditions is cylindrical collapse (similarly to spherical collapse being the most likely dynamics for 3D density fluctuations). Then the rate function of the late-time density field in concentric disks of radii \(R_i\) is given by

\[
\Psi_{	ext{cyl}}(\{\rho_i\}) = \frac{\sigma_{R_i}^2}{2} \sum_{i,j} \Xi_{k,j}(\{\tau_i\}) \bar{\tau}_i \bar{\tau}_j,
\]

where \(\sigma_{R_i}^2\) – our driving parameter – is the variance within the smallest disk, \(\Xi_{k,j}(\{\tau_i\})\) is the inverse of the covariance matrix between the linear density field inside the initial disks (before collapse) of radii \(R_k, R_j^{1/2}\) (given thanks to mass conservation in each collapsing disk), and \(\bar{\tau}_i\) are the linear density contrasts obtained through the most probable mapping between the linear and late-time density fields. This mapping is given by the 2D spherical (cylindrical) collapse for which an accurate parametrisation is given by\(^3\)

\[
\zeta(\bar{\tau}_i) = \rho_k = \left(1 - \frac{\bar{\tau}_i}{\nu}\right)^{-\nu}. \tag{24}
\]

In the spirit of previous works involving the density filtered in spherical cells, the value of \(\nu\) in this parametrisation of \(\zeta\) is chosen to be \(\nu = 1.4\) so as to reproduce the value of the tree-order skewness in cylinders as computed from perturbation theory (Uhlemann et al. 2018b).

Finally, as a straightforward consequence of the contraction principle, the rate function given by equation (23) is also the rate function of any monotonic transformation of \(\rho\), such that for the density contrast \(\delta = \rho - 1\), we have \(\Psi_{\delta}(\delta) = \Psi_{\rho}(\rho(\delta))\). Thus plugging equation (23) in equation (20) gives us the joint SCGF of concentric disks of the density field at redshift \(z\). Since we are, as shown in equation (9), interested in the density slope between two concentric disks, all that remains is to recall that its SCGF is easily expressed from its joint statistics through

\[
\varphi_{\delta_2 - \delta_1}(\lambda) = \varphi_{\delta_1, \delta_2}(-\lambda, \lambda). \tag{25}
\]

Then, the CGF can be derived and eventually the statistics of the aperture mass field can be obtained through the projection formula given by equation (9).

\section{2.5 Non-linear cumulant generating function}

Large deviation theory in the context of cosmic structure formation is strictly speaking only valid in the regime where the variance goes to zero. However a key point of its application to make useful predictions is to extrapolate these asymptotic results (in particular the CGF) to finite values of the variance. As noted in appendix A, one thus needs a prescription to compute the (co)variances for example appearing in equation (23). We use the non-linear power spectrum coming from Halofit (Takahashi et al. 2012) for computation of these covariances. As another sanity check to ensure that any possible discrepancies with the numerical simulation are not too much influenced by this choice, we also re-scale the projected CGF by the measured variance \(\sigma^2_{\text{Map,sim}}\) instead of the one computed with Halofit \(\sigma^2_{\text{Map,halo}}\), so that our resulting \(M_{\text{ap}}\) CGF contains the exact variance and all higher-order cumulants given by the non-linear collapse dynamics. The need for an exterior input of the non-linear variance along the line of sight does weaken a bit the “from first principles” quality that large-deviation frameworks in Cosmology usually exhibit but nonetheless note that such contributions are contained in the non-linear power spectrum which has focused a lot of attention in the recent years and is very reliably modelled now (Takahashi et al. 2012; Euclid Collaboration et al. 2019). Eventually, this comes with the modelling of a true non-Gaussian observable (up to a reduced shear correction but as opposed to the convergence which needs reconstruction of mass maps) which is sensitive to multiple scales at once, a property usually very useful to break degeneracy between cosmological parameters (Boyle et al. 2020).

We now have all the tools to successfully compute the non-linear cumulant generating function of the aperture mass: i) Given a non-linear prescription for the power spectrum we can compute the covariance matrix at redshift \(z\) between any two disks of radius \(R_1\) and \(R_2\)

\[
\sigma^2(R_1, R_2; z) = \frac{d^3k}{(2\pi)^3} P(k; z) W_{TH}(R_1 k_\perp) W_{TH}(R_2 k_\perp), \tag{27}
\]

where \(W_{TH}(l) = 2J_1(l)/l\) and \(J_1\) is the first Bessel function of the first kind; ii) this enables to compute the rate function (23) for any values of the densities inside each disk in a given slice along the line of sight; iii) numerically inverting the stationary condition (21) and using equations (20) and (25), we can now compute for any \(\lambda\) the CGF of the 2D density slope within each slice; iv) Using the projection formula (9) and equation (26) we finally compute the non-linear CGF of the aperture mass.

\section{2.6 Analytical cumulant generating function}

Now that we can compute the non-linear CGF of the aperture mass we would like to compute its PDF. However, the Laplace transform in equation (17) requires to have an analytical expression of the integrand, in particular the CGF, so that it can be continued in the complex plane. Unfortunately this is not the case with the described formalism, the crux of...
the matter residing in inverting the stationary condition \((21)\) for complex \(\lambda\) values given that the covariance \((27)\) is only defined numerically and that no explicit solution is known for a generic power spectrum except the simplistic case of power-law power spectra.

The solution we propose here was first used in Bernardeau & Valageas (2000) and consists in fitting an effective mapping, \(\zeta(\tau_{\text{eff}})\), between an effective un-smoothed Gaussian initial field and the aperture mass field whose PDF we want to compute. We hence re-write the aperture mass SCGF, still given by the Legendre transform of the effective rate function

\[
\varphi_{M_{\text{ap}}}(\lambda) = \lambda \zeta(\tau_{\text{eff}}) - \frac{1}{2} \tau_{\text{eff}}^2 \tag{28}
\]

with the stationary condition written as

\[
\lambda = \frac{d}{d\zeta} \frac{\tau_{\text{eff}}^2}{2} = \tau_{\text{eff}} \left( \frac{d\zeta(\tau_{\text{eff}})}{d\tau_{\text{eff}}} \right)^{-1}. \tag{29}
\]

Note that now the effective mapping

\[
\zeta(\tau_{\text{eff}}) = \sum_{k=0}^{n} \frac{\mu_k}{k!} \tau_{\text{eff}}^k, \tag{30}
\]

where \(\mu_0 = 0, \mu_1 = 1\) and the other coefficients will be fitted, makes for an easy analytic continuation of the mapping to the complex plane which in turns allows us to invert the stationary condition for complex values of \(\lambda\).

All that remains is to fit the values of the \(\mu\) coefficients. First let us notice that by definition of the Legendre transform we have

\[
\frac{d\varphi_{M_{\text{ap}}}}{d\lambda} = \zeta(\tau_{\text{eff}}), \tag{31}
\]

and thus from equation \((28)\)

\[
\frac{1}{2} \tau_{\text{eff}}^2 = \lambda \frac{d\varphi_{M_{\text{ap}}}}{d\lambda} - \varphi_{M_{\text{ap}}}(\lambda). \tag{32}
\]

Then having already computed the SCGF for real values of \(\lambda\) one can easily produce a table of both \(\zeta(\tau_{\text{eff}})\) and \(\tau_{\text{eff}}\) and finally fit the \(\mu\) coefficients. In practice, we typically choose a polynomial mapping of odd degree higher than 5 which reproduces very well the real generating function. Note that the same procedure can be applied to directly fit the extrapolated aperture mass CGF, the \(\mu\) coefficients now taking the values

\[
\mu_k^{\text{CGF}} = \varphi_{M_{\text{ap}}}^{\lambda_k}, \tag{33}
\]

and the inverse Laplace transform of the CGF in equation \((17)\) can now be performed numerically – for example using Simpson’s method along the imaginary axis – without any additional complication. We finally give in appendix D the link that can be made between the \(\mu\) coefficients and the cumulants of the field. Given that the fitted \(\mu\) values and the computed cumulants agree very well, computing cumulants to obtain those coefficients instead of fitting them might become a viable option, especially in cases where the large deviation formalism is used for filters other than top-hats and where imposing the stationary condition even to obtain the real space CGF is a numerical challenge in itself (Reinberg & Bernardeau 2018).

Figure 3 illustrates how well this approach reproduces the \(M_{\text{ap}}\) cumulant generating functions computed with large deviation theory. In particular this procedure being from a mathematical point of view strictly identical to the usual large deviation approach applied to the matter density field one-point statistics, the critical behaviour that the \((S)\)CGF exhibits along the real axis and that is the result of a change of convexity of the rate function or equivalently multiple solutions to the stationary condition, will be also present in the case of reconstructed generating functions via an effective mapping. Finally note that though the procedure we described in this section is the one that we implement to construct the aperture mass PDF, we also give some more technical comments on the approach itself for projected quantities in general in appendix B.

2.7 Aperture mass PDF

As an illustration, Fig. 4 displays the resulting prediction for an aperture mass PDF for a single source redshift at \(z_s = 1\). The opening angle is chosen to be \(\theta_1 = 15\) arcmin and the non-linear covariance is treated in three different ways, namely i) the linear prediction with a re-scaling of the obtained SCGF by the non-linear driving parameter – the variance of the field at the smallest scales – inside each redshift slice along the line of sight, ii) using the full Halofit power spectrum as input [our baseline approach] and iii) using the Euclid emulator of the non-linear power spectrum (Euclid Collaboration et al. 2019) for comparison of different non-linear prescriptions. The blue solid line shows that indeed re-scaling the SCGF by the driving parameter as in the 1-cell case does not lead to the correct non-linear variance of the aperture mass which leads to a major source of disagreement between the model and the physical PDF. This further illustrates the discussion in section 2.5 where the importance of using the full non-linear power spectrum was underlined. However, the \(M_{\text{ap}}\) PDF does not appear to be very sensitive to the precision in the modelling of the non-linear power

Overall Fig. 3 illustrates how well this approach reproduces the \(M_{\text{ap}}\) cumulant generating functions computed with large

\[
\varphi_{M_{\text{ap}}}(\lambda) = \lambda \zeta(\tau_{\text{eff}}) - \frac{1}{2} \tau_{\text{eff}}^2
\]
which illustrates that the covariances. The green and red curves are almost indistinguishable full non-linear CGF with different prescriptions for the non-linear parameter. The green and red solid lines are the PDF obtained from the linear covariance is re-scaled by the non-linear driving parameter. The green, yellow and blue PDFs are obtained fitting an effective collapse of degree 7 and we compare different prescriptions for the non-linear covariance of disks in redshift slices along the line of sight. The blue line is the traditional 1-cell approach where the CGF computed with the linear covariance is re-scaled by the non-linear driving parameter. The green and red solid lines are the PDF obtained from the full non-linear CGF with different prescriptions for the non-linear covariances. The green and red curves are almost indistinguishable which illustrates that the $M_{ap}$ PDF is not sensitive to the detailed modelling of the non-linear matter power spectrum.

Figure 4. Aperture mass PDF at $z_s = 1$, $\theta_1 = 15$ and $\theta_2 = 30$ arcmin. PDFs are obtained fitting an effective collapse of degree 7 and we compare different prescriptions for the non-linear covariance of disks in redshift slices along the line of sight. The blue line is the traditional 1-cell approach where the CGF computed with the linear covariance is re-scaled by the non-linear driving parameter. The green and red solid lines are the PDF obtained from the full non-linear CGF with different prescriptions for the non-linear covariances. The green and red curves are almost indistinguishable which illustrates that the $M_{ap}$ PDF is not sensitive to the detailed modelling of the non-linear matter power spectrum.

2.8 $M_{ap}$ PDF with a nulling strategy

One of the important issues faced by theoretical approaches that aim at describing quantities projected along the line of sight, is the mixing of both very non-linear scales not accurately probed by standard first principles perturbative approaches such as ours, and reasonably larger (quasi-linear) scales more accessible to the theory. As such usual weak-lensing statistical probes are often modelled by more phenomenological approaches such as halo models that can also take into account baryonic physics which becomes important at small scales (Mead et al. 2020), and even more so making use of numerical simulations (Schneider et al. 2019). However those simulations are not always tested in fine details, especially for higher-order non-Gaussian statistics.

Alternatively, a theoretical strategy to disentangle scales in lensing quantities known as the Bernardeau-Nishimichi-Taruya (BNT) transform or nulling strategy was proposed by Bernardeau et al. (2014c) and allows for very accurate theoretical predictions in the context of power spectrum analysis or more recently the convergence PDF (Barthelemy et al. 2020a). This nulling strategy was used very recently in Taylor et al. (2020) to remove the sensitivity to the poorly modelled small scales for the two-point cosmic shear signal, and therefore improve cosmological constraints using the Dark Energy Survey shear data. This will become even more relevant for future lensing experiments with better knowledge of redshifts.

This BNT transform can only be used in the context of a tomographic analysis of at least 3 source redshifts (or redshift bins, although not treated here) and is a linear transformation $M$ applied to the set of lensing kernels $\omega_i \equiv \omega(\chi,\chi_{s,i})$ giving rise to a new set of re-weighted kernels

$$\tilde{\omega}^j = M^j \omega_i.$$  \hfill (34)

For a set of 3 source planes labeled from $j = i - 2$ to $j = i$ arranged by ascending source redshift, it was showed in Bernardeau et al. (2014c) that $M$ must satisfy the system

$$\begin{align*}
\sum_{j=i-2}^{i} M^{ji} &= 0, \\
\sum_{j=i-2}^{i} \chi_{s,j} M^{ji} &= 0,
\end{align*}$$

which is under-constrained so that we also impose by convention $M^{ii} = 1$. The elements of $M$ can thus be computed considering sequential triplets of tomographic bins, going from the lowest to the highest redshift, such that

$$M^{i-2,i} = \frac{\chi_{i-2}(\chi_{i-1} - \chi_i)}{\chi_i(\chi_{i-2} - \chi_{i-1})},$$

$$M^{i-1,i} = \frac{\chi_{i-1}(\chi_i - \chi_{i-2})}{\chi_i(\chi_{i-2} - \chi_{i-1})}.$$  \hfill (36)

We display in Fig. 5 an example for a set of 3 source planes located at $z_s = 0.5, 1, 1.5$. The green, yellow and blue dashed lines are the kernels up to $z_s = 0.5, 1, 1.5$ respectively re-weighted by their appropriate BNT coefficients while the thick red line is the sum of the 3 re-weighted kernels. Note that the blue dashed line is also the original kernel since its BNT coefficient is set to 1. One can thus clearly see that the effect of nulling is to set to zero the contribution of all lenses below the closer plane and thus to cancel out the contribu-
Figure 5. Illustration of the effect of the BNT transform on lensing kernels. The green, yellow and blue dashed lines are the kernels up to $z_s = 0.5, 1, 1.5$ respectively re-weighted by their appropriate BNT coefficients, $M^n = [0.324, -1.324, 1]$ where $j$ is fixed and equal to 3 if the blue kernel is the third of a tomographic analysis. The thick red line is the sum of the 3 re-weighted kernel. The effect of nulling is to set to zero the contribution of all lenses below the closer plane.

3 THEORETICAL PREDICTIONS AND NUMERICAL SIMULATIONS

To make a precise assessment of the validity domain of such an approach – for instance in terms of angular scales – comparisons with numerical simulations are mandatory. It is to be stressed that past applications of the large deviation principle agreed remarkably well with numerical results derived from simulations (Bernardeau et al. 2014a; Uhlemann et al. 2018b) and we have no reason to believe it would not be the case in this specific case. Such comparisons are actually interesting in both ways as they can be used to assess the validity regime of such theories but also to validate the accuracy of the simulations which are usually not tested against non-Gaussian statistical properties. In the case of the aperture mass, we show in this section that there are issues in the measurement of its one-point PDF in state-of-the-art full sky weak-lensing numerical simulations, which prevents a precise assessment of the validity regime of our implementation.

3.1 Numerical data

We consider a set of state-of-the-art full-sky gravitational lensing simulations generated by Takahashi et al. (2017). Note that the simulations being full-sky is important in the
| $\Omega_m$ | $\Omega_A$ | $\Omega_{cdm}$ | $\Omega_b$ | $h$ | $\sigma_8$ | $n_s$ |
|---------|---------|-------------|---------|-----|---------|-----|
| 0.279   | 0.721   | 0.233       | 0.046   | 0.7 | 0.82    | 0.97|

Table 1. Cosmological parameters used throughout this paper.

comparisons to our theoretical formalism since we both need a sufficient statistics and also long wave-modes to be present in the simulated data. There, 14 boxes with side lengths of $L = 450\, {\text{Mpc}}/h$, $2L \cdots 14L$ were prepared along with 6 independent copies. The number of particles for each box was 2048$^3$, making the mass and spatial resolutions better for smaller boxes. These boxes were placed around a fixed vertex representing the observer’s position while each box was duplicated eight times and placed around the observer using periodic boundary conditions. Spherical lens shells with width of 150 Mpc$/h$ (3 per box) were then considered to trace the resulting light-ray paths from the observer to the last scattering surface. Each box was evolved in a periodic cosmological N-body simulation following the gravitational evolution of dark matter particles without baryonic processes using GADGET2. The initial conditions were generated from second-order Lagrangian perturbation theory with the initial linear power spectrum calculated using the Code for Anisotropies in the Microwave Background (CAMB, Lewis & Bridle (2002)). It was checked that the matter power spectra agreed with theoretical predictions of the revised Halofit (Takahashi et al. 2012) and ray-tracing was performed using the public code GRAYTRIX which follows the standard multiple-lens plane algorithm in spherical coordinates using the HEALPIX algorithm. The data set eventually includes full-sky convergence maps from redshifts $z = 0.05$ to 5.3 at intervals of 150 Mpc$/h$ comoving radial distance and are freely available for download\(^4\). The adopted cosmological parameters are consistent with the WMAP-9 year result and shown in Table 1. The pixelization of the full-sky maps follows the HEALPIX ring scheme with available resolutions of NSIDE = 4096, 8192 and 16384.

### 3.2 Nulled convergence power spectrum

Let us first study the agreement between the measured and computed nulled convergence power spectrum. To that end we follow the prescription of Takahashi et al. (2017) and define the power spectrum as

$$\langle 1 + (\ell/\ell_{\text{res}})^2 \rangle C_\ell^\kappa = \int_0^{\chi_{\text{m}}} d\chi \frac{\omega(\chi_s,\chi)}{\chi^2} P_s \left( k = \frac{\ell}{\chi_s} \right).$$ (38)

Here $\ell_{\text{res}} = 1.6 \times \text{NSIDE}$ and the factor in front of $C_\ell^\kappa$ accounts for the finite angular resolution of the maps. The effect of lens-shell thickness is also taken into account in Takahashi et al. (2017) by replacing the matter power spectrum by

$$P_s(k) \rightarrow \left( 1 + c_1 k^{−\alpha_1} \right) \left( 1 + c_2 k^{−\alpha_2} \right) \alpha_3 P_s(k),$$ (39)

where the additional parameters are simulation specific and equal to $c_1 = 9.5171 \times 10^{-3}$, $c_2 = 5.1543 \times 10^{-3}$, $\alpha_1 = 1.3063$, $\alpha_2 = 1.1475$ and $\alpha_3 = 0.62793$. Note that we find that those two effects – lens-shell thickness and finite angular resolution – have very little effect at the scales we are interested in. Moreover we theoretically computed for the variance at many different scales that the so-called source-plane bias (see section 3 of Takahashi et al. (2011) for details), which states that a re-weighting of the convergence pixels by the inverse magnification should be performed before any cumulative measurements, only has a sub-percent influence on the values. For reference we plot in Fig. 7 the convergence power spectrum for a source redshift at $z_s = 1.2$. There, as expected from Takahashi et al. (2017), the agreement between the theoretical model – with a Halofit matter power spectrum as input – and the measurement in the simulation is excellent.

We then show in Fig. 8 the computed and measured nulled power spectrum of the convergence for source redshifts located at $z_s = 1.2, 1.4$ and 1.6. There, a constant bias is observed at all scales and whose amplitude is directly linked to the width of the nulled convergence lensing kernel, the narrower the greater the bias. This bias can be fully explained by taking into account the discreteness of contributing lens planes in the computation of the nulled $C_\ell^\kappa$:

$$\langle 1 + (\ell/\ell_{\text{res}})^2 \rangle C_\ell^\kappa = \sum_i \Delta_i \frac{\omega(\chi_i, \chi_s)}{\chi_i^2} P_s \left( k = \frac{\ell}{\chi_i} \right),$$ (40)

where $\Delta_i$ are the width of the lens planes, $\Delta_i = 150\, {\text{Mpc}}/h$, and $\chi_i$ are the comoving distances from the observer to the centres of those planes, $\chi_s = 150(1 - 0.5)\, {\text{Mpc}}/h$. This correction becomes more relevant here than in the no-nulling case since fewer planes contribute.

As a sanity check we also computed the leading correction to the Limber approximation given in LoVerde & Ashford (2008) and did not find any difference. This is not so surprising since, though nulling kernels are somewhat narrower, the $\ell$ values for which the Limber approximation is supposedly valid are $\ell \gg \chi/\Delta \chi \approx 5$. This still adds the benefit of effectively checking that the Limber approximation is still valid for a narrow nulled lensing kernel.

Fortunately, the bias that thus appears in the naive estimation of the variance of the field is roughly constant across all scales for a given nulled lensing kernel and thus cancels out in the estimation of reduced high-order statistical quantities. We checked it for example on the reduced skewness of the aperture mass in Fig. 9 where replacing the integrations along the line of sight by summations over the lens planes only affected the value by less than half a percent. This is also clear when comparing the line of sight integrations to measurement in the simulation as shown in Fig. 10.

Finally, the reason for this constant bias can be understood in the following way: The integrand of equation (38) is the product of i) a lensing kernel term that solely depends on the redshift of contributing lenses and not on the scale $\ell$, and ii) a matter power spectrum term, that since nulling makes us probe a narrow range of both redshifts and physical scales can be approximated by $k^{n_s}$ times some redshift dependance. Now $n_s$ would obviously change for different values of $\ell$ but since the nulled lensing kernel varies very rapidly the resulting $1/x^{n_s}$ term in the integrand will merely act on its amplitude rather than its shape up to quite significant values of $n_s$. This makes the integrand of equation (38) roughly dependent on $\ell$ only through a multiplicative term and thus the error one

\(^4\) http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/
commits on the integral replacing (38) by (40) is proportional to this multiplicative term which results in a constant bias for the $C_\ell$ across all scales. This also means that one can actually predict the amplitude of the resulting bias $b$ simply by computing

$$b = \sum_i \Delta_i \omega(\chi_i, \chi_s)^2/\chi_i^2.$$

We thus recover the $\sim 6\%$ difference that we found taking into account the discreteness of the lens planes. Varying the value of $n_s$ from 0 to -10, which are both very extreme and un-realistic values since we would expect $n_s \sim -1.5$ for our scales of interest, we find that the value of $b$ only changes by $1/1000$ thus confirming our formula for the bias.

### 3.3 Filtering methods

To convolve those maps with a difference of top-hat windows of the desired angular radii and thus access simulated $M_{\text{ap}}$ statistics, we used 2 different methods that proved to give equivalent results at more than the percent precision for skewness and kurtosis and indistinguishable by eye PDFs, thus giving us confidence in the filtering schemes. Our real-space method consists in using the `query_disc` function of `healpy` to find all pixels whose centres are located within a disk centred at one specific pixel $p$ which allows to reassign the value of $p$ as being the mean of all the pixels inside the disk. Our dual-space method consists in convolving the convergence map with the appropriate filter by decomposing both the map and filter in spherical harmonics using the `map2alm` and the `beam2bl` functions of `healpy`, convolving the map and filter in this space and then going back to pixel-space via the `alm2map` function. The 2 methods were found to agree at the level of each map and for all scales which was not a given since i) the real-space procedure does not yield an exact top-hat and ii) the $a_{\text{alm}}$ transform has a non empty kernel, meaning that a randomly generated map in pixel-space, sent to $a_{\text{alm}}$-space and back might be significantly different from the original one.

### 3.4 Measured BNT $M_{\text{ap}}$ skewness

In principle, we would expect the large deviations + nulling formalism presented in this paper to perform extremely well as the BNT transform localises the lensing kernel to a finite range of redshifts and therefore physical scales thus making...
the $M_{ap}$ one-point statistics closer to the density slope in long cylinders for which large deviation formalism was already proven very efficient for example in Uhlemann et al. (2018b). Moreover this strategy applied to convergence maps was also found to be very effective (Barthelemy et al. 2020a). Also note that a very satisfying agreement between the theory and this simulation suite was already found in Fig. 6 and that we would merely like to extend this result down to smaller scales.

The difficulty that has been encountered when trying to assess the validity regime of our theoretical approach lies in the difficulty to precisely measure and assign error bars to measured $M_{ap}$ quantities. In the context of the PDF, this can be further exemplified on the measurement of the skewness, a single number but that determines the first and foremost (thanks to cumulant hierarchy) non trivial contribution to non-gaussianity in the PDF. To this purpose, we study the BNT $M_{ap}$ skewness with opening angles of $\theta_1 = 10$ arcmin and $\theta_1 = 10$ arcmin and source planes respectively located at redshift $z_s = 1.2 - 1.4 - 1.6$. For 3 different realisations of the full sky maps – except for the greatest resolution which only offers 1 realisation – we apply our filtering scheme, measure the BNT $M_{ap}$ skewness and measure error bars as standard error on the mean computed among 8 subsamples of the full sky. As shown on Fig. 9, we find that: i) the 2 different filtering schemes give similar results with most of the time much more that the percent precision; ii) different realisations at the same resolution agree very well being statistically compatible with one another as probed by the error bars and with mean values very close to each other; iii) there is a visible shift of the value with increasing resolution to the point that the lowest and highest resolution – though coming from the same realisation – do not seem compatible, which also does not seem to come from resolution itself since degrading the map by hand to a lower resolution, filtering it and then re-measuring the skewness leads to very similar results; iv) none of the measured values agree with the theoretical prediction ($\lesssim 10\%$ relative difference). Note that to some extent, those observations are still valid in the case presented previously in Fig. 6 but smaller. There, the PDF was measured from one single realisation of the map at the lowest resolution and proved to perform very well.

Additionally, note that the discrepancy between the measured $M_{ap}$ skewness for different resolutions – but still the same realisation – of the nulled convergence map is mainly sourced by the fact that the measured $M_{ap}$ variance varies between resolutions.

A more careful examination of the BNT $M_{ap}$ skewness value can also be performed decomposing it as cross-cumulants of the BNT convergence field filtered at 2 different scales. Following equation (14) it is written as

$$S_{3,M_{ap}} = \frac{\langle M_{ap}^3 \rangle_c}{\sigma_{M_{ap}}^3} \quad (42)$$

$$= \frac{\langle \kappa_3^2 \rangle_c - 3\langle \kappa_1^2 \kappa_2 \rangle_c + 3\langle \kappa_1 \kappa_2^2 \rangle_c}{\sigma_{M_{ap}}^3}. \quad (43)$$

We then measured each cross-cumulants in the first realisation at the lowest and medium resolutions and compared them to their respective theoretical predictions as shown in Fig. 10. The agreement for those cross-cumulants is found to be excellent (to the percent precision and very close to the 1-$\sigma$ error bar), allowing us to extend the results obtained on the 1-point convergence to a multi-scale analysis. Let us notice that in eq. (43) numbers of similar orders of magnitude subtract each other which in turn tends to decrease the precision of the prediction of the BNT $M_{ap}$ skewness since absolute difference tend to stay the same while the subtraction of 2 numbers of similar magnitude results in a smaller number which finally increases the relative difference between the theory and the numerical simulation. We are thus looking at some very subtle effects in the convergence field itself which tend to have a significant impact on the quantities we are interested in. This could be the sign of higher order perturbative corrections (loop corrections) but given that observables similar to the BNT $M_{ap}$ such as densities in cylinders were already studied and found to be accurate at tree-order (Uhlemann et al. 2018b), a more likely explanation is to be found on the simulation side. Indeed subtle numerical artefacts could show up in the BNT $M_{ap}$ such as resolution or discretisation effects, convergence of the N-body algorithm etc. In fact, since the relative incoherences observed on the skewness when increasing the maps resolution are not at all seen on the convergence field itself, it could indeed be possible that small errors in the simulation are amplified when looking at $M_{ap}$ statistics.

Finally note that we also tried to perform – sometimes successfully – comparison of our theoretical formalism to other numerical suites. However other issues such as the evaluation of the degree of independence between realisations of randomised lines of sight in replicates of a small N-body box, and other simulation-specific issues prevented us from doing any more refined comparison to available ray-tracing simulations. There is therefore a dire need for future code comparisons, validations and improvements in the line of Hilbert et al. (2020) that specifically target higher-order statistics. Such future developments could then allow to test in more details the validity regime of our theoretical model for the BNT $M_{ap}$ statistics. In any case, exploitation of the $M_{ap}$ PDF as an observable should not, in light of those findings, be done assuming that numerical simulations are sufficiently accurate.
and physical formalism just like ours should be considered to at least test the different numerical schemes in cosmological analysis of non-Gaussian statistics.

3.5 Resolution effects

Before concluding, let us mention two additional tests that were done on the theory side to try to mend the discrepancy between the theory and the simulation. We separate those since they are not tied at all to the comparison to simulations in general. First, we tried to introduce a $L$-cut due to the resolution in the computation of the skewness itself which is done by introducing the corresponding $k$-cut at each slice along the line of sight thus modifying the variance and its derivatives in equation (C8). Second, we considered the effect of having a non-exact top-hat filter in appendix E choosing our $L$ parameter – typically $L = 20$ for the lowest resolution – to smooth our filter over 2 pixels. Those two prescriptions were shown to have non significant impact on the skewness and therefore are unlikely explanations for the observed differences.

4 SHAPE NOISE

Since the weak-lensing aperture mass is obtained from cosmic shear measurements by means of the measured shape of galaxies which themselves are intrinsically elliptical, the observed shear is the sum of contributions from weak-lensing and the intrinsic galactic ellipticities. Shape noise is caused by the variance of the intrinsic ellipticity, which is the dominant source of noise in shear measurements and impacts the convergence field with a random noise that can be modelled by a Gaussian with zero mean and variance $\sigma_{SN}^2$. To estimate the variance of the shape noise distribution in a Euclid-like configuration, we assume

$$\sigma_{SN}^2 = \sigma^2 N_{bin}/(n_{ga}, \Omega_0),$$

where $\sigma^2 = 0.3$ is the Gaussian noise on the determination of the lensing-induced ellipticity of each galaxy, $\Omega_0$ is a solid angle in units of arcmin$^{-2}$ – the area of a pixel – $n_{ga} = 30$ arcmin$^{-2}$ is the normalised mean number of observed galaxies, and $N_{bin} = 10$ is the number of equally populated redshift bins in the survey (Deshpande et al. 2020). The resulting Gaussian noise that this procedure thus induces on the constructed $M_{ap}$ field is then given by

$$\sigma_{SN,M_{ap}}^2 = \frac{\alpha \sigma^2 N_{bin}}{\pi n_{ga}} \left( \frac{1}{\partial \theta^1} - \frac{1}{\partial \theta^2} \right),$$

where $\alpha = 1$ for regular $M_{ap}$ and $\alpha = \sum_{i=1}^{2} (M^2)^2$ for BNT $M_{ap}$. We would then obtain the resulting noisy theoretical $M_{ap}$ PDF by convolving the previously computed PDFs with a zero-mean Gaussian of the appropriate variance

$$P_{SN}(M_{ap}) = \frac{1}{\sigma_{SN,M_{ap}} \sqrt{2\pi}} \int_{-\infty}^{+\infty} dM_{ap}P(M_{ap}) \exp \left( \frac{(M_{ap} - \hat{M}_{ap})^2}{2\sigma_{SN,M_{ap}}^2} \right).$$

As an illustration of the effect of shape noise on the $M_{ap}$ PDF we show in Fig. 11 the relative difference between the noisy PDFs $P_{SN}(M_{ap})$ and a Gaussian of the same mean and total variance (noise + signal). This is done by adding Gaussian noise to each pixels of the simulated convergence maps of 10 realisations and at redshifts mimicking a Euclid binning, combining them so as to get nulled maps, measuring the resulting $M_{ap}$ PDFs and computing the standard deviation between the 10 realisations in each bin of the $M_{ap}$ PDF as an estimate of the cosmic variance. We fix the opening angle to $\theta_2 = 2\theta_1 = 30$ arcmin and show both the residuals of the noisy $M_{ap}$ PDF with respect to a Gaussian of the same variance for a source redshift $z_s = 1.2$ and nulled bins along the line of sight up to $z_s = 1.2$.

We observe that a part of the non-Gaussian signal still remains in the regular $M_{ap}$ fields while the signal of a single bin of the BNT $M_{ap}$ is so noisy that we only observe the zero-mean Gaussian of variance $\sigma_{SN,M_{ap}}^2$. This is explained by the fact that the amplitude of the BNT $M_{ap}$ signal is by construction way smaller than that of regular $M_{ap}$ since it boils down to reducing the number of lenses contributing to the effect by localising the lensing kernel at specific physical scales and also by the fact that the noise itself is increased. Though this could thus seem like the end of the line for the nulling strategy, this is fortunately not yet the case. Indeed, at the level of a tomographic analysis where multiple source redshifts are considered, the BNT strategy is a simple reorganisation of the signal that, similarly to a principle component analysis, drastically diminishes the redundancy of the information that is present in each map as a result of scale mixing. Thus no signal is lost and though the noise is seemingly increasing in each nulled bin, a joint analysis of all bins should enable to recover all the information while having increased our ability to theoretically probe this information. Remains to write down the formalism for the joint PDF of all $M_{ap}$ nulled bin which was hinted in Barthelemy et al. (2020a) but is left for future work. On the other hand, looking at the regular $M_{ap}$ PDF at $z_s = 1.2$ and seeing that some non-Gaussianities remain can be misleading since i) a part of the signal comes from scales that are not well modelled which prevents us from extracting cosmological information out of it (as shown with the residual at $z_s = 0.57$) and ii) the signal in each redshift bin is very redundant and thus a false sense of accumulating.

Figure 10. Reduced cross-cumulants of the nulled convergence field appearing in equation (43). Points with error bars are measured in the simulation’s first realisation and solid lines represent the corresponding theoretical predictions. Even though the values tend to be quite sensitive to the resolution, a sub-percent agreement with the theoretical predictions is found.

MNRA 000, 1–20 (2020)
emulators (Euclid Collaboration et al. 2019). All subsequent attention in the recent years with very reliable models such as an input. Fortunately, this issue has received a lot of at-

tion between 10 realisations of the simulation with added noise to improve agreement between theory and simulations since this issue is not observed at the level of the convergence field and since the aperture mass enhances subtle non-linear features of the convergence field – see Fig. 10 and equation (43) – the measured values are to be handled with caution and more numerical tests should be performed. Generally speaking, higher-order weak-lensing statistics are not tested in fine details in numerical simulations and there is a real need to provide theoretical prescriptions that enable those tests. We hope that the present formalism falls in this line.

Many recent works started the daunting task of evaluating the accuracy of weak-lensing simulations for non-Gaussian statistics: (Hilbert et al. 2020). Matilla et al. (2020) studied the impact of thickness of the lens planes used to build past light-cones and the mass resolution of the underlying N-Body simulation and thus proposed guiding lines for the design of future numerical suites. Another issue is the pseudo-independence of lensing maps generating from randomisation of the line of sight through replication of a small but better resolved N-body box. Petri et al. (2016) for example suggested that maps generated from 1 or 2 realisations of the same box could be considered independent but those results, to our knowledge, were not replicated nor extended to the case of the convergence PDF or moments and even less for the nulling strategy, perform extremely well with no apparent deviation from the simulation given the estimated error bars, see Fig. 6, at least for sufficient large scales and redshifts. However more conclusive tests could not be performed for smaller scales due to some imprecision and inconsistencies that we found in the numerical suite like an inconsistent shift of the aperture mass skewness when increasing the map resolution. Since this issue is not observed at the level of the convergence field and since the aperture mass enhances subtle non-linear features of the convergence field – see Fig. 10 and equation (43) – the measured values are to be handled with caution and more numerical tests should be performed. Generally speaking, higher-order weak-lensing statistics are not tested in fine details in numerical simulations and there is a real need to provide theoretical prescriptions that enable those tests. We hope that the present formalism falls in this line.

Many recent works started the daunting task of evaluating the accuracy of weak-lensing simulations for non-Gaussian statistics: (Hilbert et al. 2020). Matilla et al. (2020) studied the impact of thickness of the lens planes used to build past light-cones and the mass resolution of the underlying N-Body simulation and thus proposed guiding lines for the design of future numerical suites. Another issue is the pseudo-independence of lensing maps generating from randomisation of the line of sight through replication of a small but better resolved N-body box. Petri et al. (2016) for example suggested that maps generated from 1 or 2 realisations of the same box could be considered independent but those results, to our knowledge, were not replicated nor extended to the case of the convergence PDF or moments and even less for the aperture mass. All of this suggests that indeed more work is needed if we are to extract subtle non-linear effects from weak-lensing simulated maps.

On a more theoretical side and for the specific case of the aperture mass, the fact that it enhances subtle non-linear features of the convergence field would call for evaluation of beyond cylindrical collapse contributions of the non-linear dynamics of the underlying density field, that is loop corrections in perturbation theory. This is not the main path chosen to improve agreement between theory and simulations since the BNT aperture mass is somewhat close to the density field filtered in long cylinders for which the cylindrical collapse dynamics were shown to perform extremely well (Uhlmann et al. 2018b), but this is nonetheless an interesting work that should be carried out in the future.

Future theoretical work on the aperture mass PDF or moments could also include general relativistic corrections to the non-linear Newtonian dynamics of the underlying matter density field. Such effects were already taken into con-

5 DISCUSSION AND CONCLUSION

In this paper we have presented a formalism for the one-point statistics of the aperture mass seen as a difference of top-hat filters applied to the convergence field. The geometry and time-evolution within the light-cone is taken into account relying on the fact that the correlations of the underlying matter density field along the line of sight are negligible compared to transverse directions, which leads to redshift slices being treated as statistically independent. Within each redshift slice, the most likely non-linear dynamics of the matter density field filtered in concentric disks is on average and in the limit of small variance well approximated by the cylindrical collapse, a fact which makes it possible to treat the filtered field as a random variable satisfying a large deviation principle. The large deviation formalism then allows us to compute the joint CGF of the density filtered in disks of two different radii, thus the CGF of the density slope and finally the one-point statistics of the aperture mass $M_{\text{ap}}$ through projection effect. The formalism in itself only provides means to compute the $M_{\text{ap}}$ CGF and PDF if a prescription for the non-linear matter power spectrum is used as an input. Fortunately, this issue has received a lot of attention in the recent years with very reliable models such as emulators (Euclid Collaboration et al. 2019). All subsequent non-Gaussian information could come out of it while it is effectively not as much the case as one could imagine. The joint-analysis of BNT transformed redshift bins, where shape noise is properly taken into account and model, is thus the correct way to perform a tomographic analysis. We leave for a further work a precise analysis of its performance.

Figure 11. Residuals of the noisy $M_{\text{ap}}$ PDFs with respect to a Gaussian of the same variance. The opening angle is $\theta_1 = 15$ and $\theta_2 = 30$ arcmin. The regular $M_{\text{ap}}$ is taken at source redshift $z_s = 1$ and the BNT $M_{\text{ap}}$ has source redshifts located at $z_s = 1.2$, 1.4 & 1.6. Error bars are taken as the standard deviation between 10 realisations of the simulation with added noise to estimate the cosmic variance.

Information could come out of it while it is effectively not as much the case as one could imagine. The joint-analysis of BNT transformed redshift bins, where shape noise is properly taken into account and model, is thus the correct way to perform a tomographic analysis. We leave for a further work a precise analysis of its performance.

Figure 11. Residuals of the noisy $M_{\text{ap}}$ PDFs with respect to a Gaussian of the same variance. The opening angle is $\theta_1 = 15$ and $\theta_2 = 30$ arcmin. The regular $M_{\text{ap}}$ is taken at source redshift $z_s = 1$ and the BNT $M_{\text{ap}}$ has source redshifts located at $z_s = 1.2$, 1.4 & 1.6. Error bars are taken as the standard deviation between 10 realisations of the simulation with added noise to estimate the cosmic variance.
sideration in Lepori et al. (2020) on the convergence PDF by means of ray-tracing through a relativistic N-body simulation but a detailed comparison with the Newtonian PDF was not performed. Nevertheless, should this be done and a significant effect be found, those type of corrections could be computed within the post-Friedmann formalism for which a re-definition of the usual weak-lensing fields was recently performed in Gressel et al. (2019). Another crucial correction to be taken into account comes from the fact that the observed field, the lensing-induced ellipticities of galaxies, correspond to the reduced shear \( \gamma/(1 + \kappa) \) rather than the shear itself. Fortunately, the leading correction to the observed \( M_{ap} \) skewness that arises can be computed with perturbation theory as was done in Schneider et al. (1998) and in appendix F and shown to be a few percent of the skewness value. This can also be performed at the level of the PDF of the density slope within a redshift slice by means of large deviation theory (Reimberg & Bernardeau 2018). Hence in principle, computing the observed \( M_{ap} \) PDF by fully taking into account the geometry of the light-cone is straightforward, but since it is numerically involved and that there are already challenges to compare the \( M_{ap} \) PDF to simulations we leave it to future work.

Still based on Reimberg & Bernardeau (2018), we also showed in appendix E how to account for compensated filters that are different from a difference of top-hat windows. Overall we show that even filters that deviate significantly from the top-hat lead to only a few percents difference on the skewness which states that our formalism could be readily implemented even with smoother filtering schemes, keeping the relative theoretical simplicity of top-hats and being able to model the systematic deviation coming from this approximation. This could be particularly useful knowing that some compensated filters do offer better cosmological constraints than others as is discussed for example in Lin et al. (2016).

Finally, note that to be realistic, one would also need to account for the fact that the aperture mass is measured not from a single source redshift but from a given source galaxy distribution \( n_s(z_s) \). This can be readily done in our formalism. Indeed, when aiming to predict the weak-lensing aperture mass measured from \( n_s(z_s) \), one can simply replace the lensing kernel \( \omega \) by

\[
\omega_{ns}(\chi) = \frac{3 \Omega_m H_0^2}{2 c^2} \int_0^\infty d\chi_s \frac{D(\chi) D(\chi_s - \chi)}{D(\chi_s) a(\chi)} n_s(z_s) \frac{dz_s}{d\chi_s}, \tag{47}
\]

and only the implementation of the nulling strategy would be slightly different (Bernardeau et al. 2014c).

ACKNOWLEDGEMENTS

This work is partially supported by the SPHERES grant ANR-18-CE31-0009 of the French Agence Nationale de la Recherche and by Fondation MERAC. AB’s work is supported by a fellowship from CNES. This work has made use of the Horizon Cluster hosted by Institut d’Astrophysique de Paris. We thank Stéphane Rouberol for running smoothly this cluster for us, Ken Osato for pointing us to the simulation we used in this paper as well as discussing its issues with us and Simon Prunet for fruitful discussions. AB thanks the Institut de Physique Théorique for hosting several visits during the completion of this work. We also thank Aoife Petrie, Oliver Friedrich, Cora Uhlemann, Martin Kilbinger and Raphaël Gavazzi for fruitful discussions.

DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material.

References

Bard D., Kratochvil J. M., Dawson W., 2016, ApJ, 819, 158
Barthelemy A., Codis S., Uhlemann C., Bernardeau F., Gavazzi R., 2020a, MNRAS, 492, 3420
Barthelemy A., Codis S., Bernardeau F., 2020b, MNRAS, 494, 3369
Bangh C. M., Gaztanaga E., Efstathiou G., 1995, MNRAS, 274, 1049
Bernardeau F., 1995, A&A, 301, 309
Bernardeau F., Reimberg P., 2016, Phys. Rev. D, 94, 063520
Bernardeau F., Valageas P., 2000, A&A, 364, 1
Bernardeau F., van Waerbeke L., Mellier Y., 1997, A&A, 322, 1
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Bernardeau F., Pichon C., Codis S., 2014a, Phys. Rev. D, 90, 103519
Bernardeau F., Pichon C., Codis S., 2014b, Phys. Rev. D, 90, 103519
Bernardeau F., Nishimichi T., Taruya A., 2014c, MNRAS, 445, 1526
Boyle A., Uhlemann C., Barthelemy A., Friedrich O., Codis S., Bernardeau F., 2020, In prep
Deshpande A. C., et al., 2020, A&A, 636, A95
Euclid Collaboration et al., 2019, MNRAS, 484, 5509
Gavrilidis P., Athanassoulis G., 2009, Journal of Computational and Applied Mathematics, 229, 7
Gressel H. A., Bonvivin C., Bruni M., Bacon D., 2019, J. Cosmology Astropart. Phys., 2019, 045
Hilbert S., et al., 2020, MNRAS, 493, 305
Ivezić Ž., et al., 2019, ApJ, 873, 111
Kacprzak T., et al., 2016, MNRAS, 463, 3653
Kaiser N., 1992, ApJ, 388, 272
Kaiser N., 1995, ApJ, 439, L1
Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901
Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193
Lepori F., Adamek J., Durrer R., Clarkson C., Coates L., 2020, MNRAS, 493, 2078
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Lin C.-A., Kilbinger M., Pires S., 2016, A&A, 593, A88
LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506
Martinet N., et al., 2018, MNRAS, 474, 712
Martinet N., Harnois-Dérap J., Julio E., Schneider P., 2020, arXiv e-prints, p. arXiv:2010.07476
Matilla J. M. Z., Waterval S., Haiman Z., 2020, AJ, 159, 284
Mead A., Brieden S., Tröster T., Heymans C., 2020, arXiv e-prints, p. arXiv:2009.01858
Mellier Y., 1999, Annual Review of Astronomy and Astrophysics, 37, 127
Munshi D., Valageas P., Barber A. J., 2004, MNRAS, 350, 77
Patton K., Blazek J., Honscheid K., Huff E., Melchior P., Ross A. J., Suchyta E., 2017, Monthly Notices of the Royal Astronomical Society, 472, 439
Peebles P. J. E., 1980, The large-scale structure of the universe
Petri A., Haiman Z., May M., 2016, Phys. Rev. D, 93, 063524
Petri A., Haiman Z., May M., 2017, Phys. Rev. D, 95, 123503
Planck Collaboration et al., 2020, A&A, 641, A1
APPENDIX A: NON-LINEAR COVARIANCE ALONG THE LINE OF SIGHT

As a consequence of extending the results of large deviation theory to finite, non-zero values of the variance, one needs a prescription to compute the (co)variances for example appearing in equation (23). Past results (for instance Uhlemann et al. (2018a)) on the 3D density field relied on modelling the non-linear variance by re-scaling the linear value

\[ \sigma_{nl}(Rp^{1/2}) = \frac{\sigma_{nl}(R)}{\sigma_{l}(R)} \sigma_{l}(Rp^{1/2}), \]  

(A1)

where \( \sigma_{nl}(R) \) was acting as the driving parameter taken in its non-zero, finite value not predicted by the theory and thus left as a free parameter and measured in data or numerical simulations, or computed with some non-linear prescription for the power spectrum.

This construction ensures that the reduced cumulants of the density field are exactly those obtained through standard tree-order PT and is also what one should typically use for projection of the density field, for example the convergence field as was shown in Barthelemy et al. (2020a). However, for the case of the joint statistics of the density field at different scales the choice of a unique driving parameter prevents us from imposing all the correct quadratic contributions in the CGF which is particularly problematic for the 1-point PDF of the aperture mass. Note indeed that sums of random variables satisfying a large deviation principle do not necessarily satisfy the same principle. As a consequence, we in this work choose to model the non-linear covariance, not by a re-scaling by the driving parameter which thus do not matter in the final expression of the CGF, but by a full non-linear prescription coming in our case from Halofit. This ensures that all quadratic contributions in the CGF are correctly modelled and modifies the standard tree-order PT results keeping their functional but using a non-linear power spectrum where the linear ones usually appears. One would then need to compare these results to high-order loop calculations to check the meaning of these corrections, this is left for future work.

APPENDIX B: TECHNICAL COMMENTS ON THE EFFECTIVE MAPPING APPROACH TO THE APERTURE MASS (S)CGF

The procedure described in section 2.6 for the aperture mass CGF – i.e a projected quantity – is from a mathematical point of view strictly identical to the construction of the density slope CGF in each redshift slice along the line of sight, solely changing the physics (cylindrical collapse VS effective collapse) and the initial conditions (smoothed VS un-smoothed Gaussian field). This in particular means that the effective procedure fails, at least on paper, to convey some of the information that the large-deviation approach to the aperture mass CGF does encode. Though this is not reflected in the successive cumulants of the field which are well reproduced by the effective approach, this is the case in the tails of the PDF or equivalently in the CGF near the critical points. Indeed it can be shown (Bernardeau et al. 2014a) by expanding the stationary condition, be it (21) or (29), near a critical point \( \lambda_c \) that the density CGF behaves like

\[ \phi_\rho(\lambda) \sim (\lambda - \lambda_c)^{3/2}, \]

which leads to the PDF exhibiting an exponential cut-off in \( P_\rho(\rho) \sim \exp(-\lambda_c \rho)/\rho^{3/2} \). As for projected CGFs, projection effects do modify their asymptotic behaviour near critical points since they now behave like – see for instance appendix A of Bernardeau & Valageas (2000) – \( \phi_{proj}(\lambda) \sim (\lambda - \lambda_c)^2 \log(\lambda - \lambda_c) \) which obviously changes the exponential cut-off of the PDF. Thus a more tedious but more appropriate effective approach to the analytical aperture mass CGF would rather be to fit an effective mapping at the level of each redshift slice along the line of sight. Fortunately, we did not in practice find any significant difference in the PDF tails between fitting an effective collapse at each slice or directly on the projected CGF, which is explained by the fact that the asymptotic result is reached rather far in the tails (Bernardeau et al. 2014b) and thus the main contributions to the shape of the PDF come from the few first cumulants.

A similar remark can be made regarding the degree of the effective mapping that directly influences the positions of the critical points along the real axis. Indeed we saw in the previous paragraph that the asymptotic behaviour of the PDF does change with their positions. However, we again find that this is not in practice a crucial issue and find that a polynomial mapping of degree 5 well describes the numerical generating functions.

Finally, we find it interesting to explicit that effectively, only the knowledge of the first few cumulants allows us to recover the entire PDF. This is as we saw enforced by the formal construction obtained through assumption that the \( M_{ap} \) satisfies – this is only true with the effective mapping – a large deviation principle, and also by the fact that we impose our rate function to be convex through usage of the Legendre rather than Legendre-Fenchel transform, which effectively acts as if the rate function was linear beyond the critical points. The unchanging convexity of the rate function imposes that the PDF is uni-modal. This can be for example put into perspective with works in the field of mathematics, see for example Gavrilidis & Athanassoulis (2009), that present, in general, how much of the PDF can be recovered just from the information of a certain number of moments. It turns out that uni-modal PDFs can be very well approximated by a formal and general reconstruction coming from
APPENDIX C: APERTURE MASS CUMULANTS AT TREE ORDER

We derive in this section the expression of the first cumulants of the aperture mass before integration along the line of sight. This derivation is greatly facilitated by use of the large deviation principle and is strictly equivalent to perturbation theory up to the first non-trivial term (i.e. at tree-order).

Let us first derive the expression for joint cumulants of the density field. The exact spherical collapse mapping can be written as

\[ \rho(\tau) = 1 + \delta(\tau) = \sum_{k=0}^{\infty} \frac{\nu_k}{k!} \tau^k \]  

(C1)

where \( \tau \) is the mean linear density contrast given through the most probable mapping between the linear and late-time density field, \( \delta \) its non-linear counterpart, and where \( \nu_k \) are the spherically averaged perturbation theory kernels (Bernardeau et al. 2002)

\[ \nu_k = k! \int d\Omega_1 \ldots d\Omega_k F_i(k_1, \ldots, k_k). \]  

(C2)

Then the joint rate function of the density field filtered in two disks of radii \( R_1 \) and \( R_2 \) is

\[ \psi(\rho_1, \rho_2) = \frac{1}{2} \sum_{k,j} \xi_{kj}(\rho_1^{1/2} R_1, \rho_2^{1/2} R_2) \tau(\rho_k) \tau(\rho_j), \]  

(C3)

and its associated CGF is

\[ \begin{aligned} \phi(\lambda_1, \lambda_2) &= \lambda_1 \rho_1 + \lambda_2 \rho_2 - \psi(\rho_1, \rho_2) \\ \lambda_1 &= \frac{\partial}{\partial \rho_1} \psi(\rho_1, \rho_2) \\ \lambda_2 &= \frac{\partial}{\partial \rho_2} \psi(\rho_1, \rho_2). \end{aligned} \]  

(C4)

This allows to define the two operators

\[ \begin{aligned} D_1 &= \frac{\partial}{\partial \lambda_1} = \frac{\psi_{22}}{\det \psi_{ij}} \frac{\partial}{\partial \rho_1} - \frac{\psi_{12}}{\det \psi_{ij}} \frac{\partial}{\partial \rho_2} \\ D_2 &= \frac{\partial}{\partial \lambda_2} = \frac{\psi_{11}}{\det \psi_{ij}} \frac{\partial}{\partial \rho_1} - \frac{\psi_{12}}{\det \psi_{ij}} \frac{\partial}{\partial \rho_2} \end{aligned} \]  

(C5)

which leads to

\[ \langle \rho_1^{1+p} \rho_2^{1+q} \rangle_c = \left. D_1^p D_2^q \phi(\lambda_1, \lambda_2) \right|_{\lambda_1=0, \lambda_2=0}. \]  

(C6)

The same can be done in the simpler case of the 1-cell CGF and one gets

\[ \langle \rho^{2+p} \rangle_c = \left. \frac{d \phi(\lambda)}{d \lambda^{2+p}} \right|_{\lambda=0} = \left( \frac{1}{\psi''(\rho)} \right)^p \left( \frac{1}{\psi''(\rho)} \right)^p \right|_{\tau=0}. \]  

(C7)

Then combining equations (C6),(C7) and (14) one gets up to the integration along the line of sight

\[ \langle M_{\text{ap}}^2 \rangle_c = \sigma_{M_{\text{ap}}}^2 \left[ \sigma_{M_{\text{ap}}}^2 (R_2) \left[ 3 \nu_2 + \frac{3 \log \sigma_{M_{\text{ap}}}^2}{2 \log R_2} \right] \right]^{1/2} \]  

where

\[ \sigma_{M_{\text{ap}}}^2 = \sigma^2(R_1) + \sigma^2(R_2) - 2 \sigma^2(R_1, R_2). \]  

(C9)

Other cumulants can of course also be obtained in a similar fashion although their algebraic expression might become more and more complicated. The procedure can nonetheless be implemented in a symbolic calculus software and be used there.

APPENDIX D: EFFECTIVE MAPPING AND APERTURE MASS CUMULANTS

We here compute the relation between the \( \mu \) coefficients of the effective mapping

\[ \zeta(\tau_{\text{eff}}) = \sum_{k=0}^{n} \frac{\mu_k}{k!} \tau_{\text{eff}}^k, \]  

(D1)

and the cumulants of the aperture mass whose SCGF is given by

\[ \phi_{M_{\text{ap}}}(\lambda) = \lambda(\tau_{\text{eff}}) - \frac{1}{2} \tau_{\text{eff}}^2, \]  

(D2)

with the stationary condition written as

\[ \lambda = \frac{d \tau_{\text{eff}}^2}{d \zeta^2} = \tau_{\text{eff}} \frac{d \zeta(\tau_{\text{eff}})}{d \tau_{\text{eff}}}^{-1}. \]  

(D3)

Applying the same procedure leading to equation (C7) we arrive at

\[ S_{8,M_{\text{ap}}} = 3 \mu_2, \]  

(D4)

\[ S_{4,M_{\text{ap}}} = 12 \mu_2^2 + 4 \mu_3, \]  

(D5)

\[ S_{6,M_{\text{ap}}} = 60 \mu_2^2 + 60 \mu_2 \mu_3 + 5 \mu_4, \]  

(D6)

\[ S_{6,M_{\text{ap}}} = 360 \mu_2^3 + 720 \mu_2^2 \mu_3 + 90 \mu_3^2 + 120 \mu_2 \mu_4 + 6 \mu_5. \]  

(D7)

These relations can be inverted and one finally arrives at

\[ \mu_2 = \frac{S_8}{3}, \]  

(D8)

\[ \mu_3 = \frac{-4 S_4 + 3 S_8}{12}, \]  

(D9)

\[ \mu_4 = \frac{40 S_4^2 - 45 S_8 S_4 + 9 S_8}{45}, \]  

(D10)

\[ \mu_5 = \frac{-560 S_4^3 + 840 S_4^2 S_4 - 135 S_8^2 - 192 S_8 S_4 + 24 S_8}{144}. \]  

(D11)

APPENDIX E: SENSITIVITY TO THE FILTERING SCHEME

One issue of top-hat filtering is its 'sharpness' (non differentiability at the radius) which could make its precise imple-
normalisation of the filter defined by equation (E1) is built so that the departure from an actual top-hat in the limit where \( \kappa \rightarrow 0 \) is i) centred around the radius and ii) a fixed percentage of its value for a given \( L \). Finally note that \( \kappa \) is a parameter that influences the sharpness of the filter – we recover a top-hat from the exact top-hat for this \( L \) is roughly \( \pm 10\% \) around the radius.

For our definition of the aperture mass as the difference of the convergence smooth-top-hat filtered at two different scales,

\[
\hat{U}_{M_{ap}}(\vartheta) = \frac{4\theta^2 e^{L^2}L^3}{\sqrt{\pi}e^{L^2}(2L^2 + 1)} \left( \frac{\exp(-L^2(\theta_2 - \vartheta)^2/\theta_2^2)}{\theta_2^2} - \frac{\exp(-L^2(\theta_1 - \vartheta)^2/\theta_1^2)}{\theta_1^2} \right)
\]

is shown in Fig. E2. Note that \( \hat{U}_{M_{ap}} \) would be a difference of two Dirac delta functions if we were using regular top-hats as is done in the main text. For simplicity we also use the same value of \( L \) for the two different scales which leads to broader \( \hat{U}_{M_{ap}} \) as \( \theta_1 \) increases though this might not be an important issue for radii relatively close one to another.

We now need to express the SCGF of the density in each slice along the line of sight. This accounts to write the continuous limit of equations (19) and (23) re-written as

\[
\varphi(\lambda) = \sup_{\{\rho_i\}} \left[ \lambda \sum_{i=1}^{N} \tilde{U}_i \rho_i - \frac{\sigma^2}{2} \sum_{k,j}^{N} \xi_{kj}(\{\tau_i\}) \gamma_k \gamma_j \right].
\]

This is done by writing

\[
\sum_{i=1}^{N} \tilde{U}_i (R_i) \rho_i \rightarrow \int dR \hat{U}(R) \xi(\tau_R)
\]

with \( \rho_i = \zeta(\tau_i) \) the cylindrical (2D spherical) collapse, \( R \) the smoothing radius of the non-linear density and \( r \) its value for the linear field expressed through mass-conservation as

\[
r = R \rho^{1/2}.
\]

We also need a continuous limit to the cross-correlation matrix \( \Xi \) which is obtained assuming the existence of an object \( \xi(r', r'') \) defined by

\[
\int dr' \sigma^2(\rho, r') \xi(r', r'') = \delta_D(r - r'').
\]

Therefore the continuous limit to the SCGF can be written
as

$$\varphi(\lambda) = \sup_\tau \left[ \lambda \int \frac{dR}{dr} \hat{U} [R(r)] \zeta(\tau) - \frac{\sigma_{\text{slice}}^2}{2} \int dr dr' \tau \tau' \xi(r, r') \right],$$  \hspace{1cm} \text{(E12)}

and $\sigma_{\text{slice}}^2 = \int dr dr' \sigma^2(r, r') \hat{U}(r) \hat{U}(r')$. Note that the practical implementation of this extremaization problem is not straightforward as shown in Reimberg & Bernardeau (2018).

Thus expanding the values of $\varphi(\lambda)$ around 0 leads to the expression of the reduced cumulants. For the skewness in each slice one gets

$$\varphi_{\text{slice}}^3 = 3 \nu_2 \left[ \int \frac{d\pi W}{d\pi} \hat{U}(x) \Sigma^2(x) + 3 \int \frac{d\pi W}{d\pi} \hat{U}(x) \Sigma'(x) \right].$$ \hspace{1cm} \text{(E13)}

where the coefficient in front of the second term is actually 6 over the dimension of the collapse, $\nu_2$ is the usual spherical collapse coefficient and we have

$$\Sigma(x) = \int d\pi W^2(x, y) \hat{U}(y).$$ \hspace{1cm} \text{(E14)}

Now combining the result in equation (E13) and the projection formula (9) one can get to the aperture mass skewness taking into account the geometry of the light-cone as well as the smoothed top-hat filter (E1). For a nulled aperture mass field with source redshifts located at $z_s = 1.2 - 1.4 - 1.65$ and smoothing angles $\theta_1 = 10$ and $\theta_2 = 20$ arcsim we both compute the variance and the reduced skewness using an exact top-hat window and its smoothed version for different values of $L$. We also use an Halofit power-spectrum. The relative difference between the two is displayed in Fig. E3 where on can see that the values of both the variance and skewness depend very weakly on the value of $L$. This is overall good news since approximate filtering schemes could thus be considered with still relatively good validity of our theoretical modelling.

APPENDIX F: SECOND ORDER CORRECTIONS

F1 Post-born corrections

Throughout this work we modelled the aperture mass as a weighted integral over the underlying non-linear density field, therefore assuming independent lenses and following light rays along un-perturbed lines of sight in a $\Lambda$CDM universe. Hence two ingredients were neglected: i) couplings between the lenses which state that the combination of lenses in geometrical optics is not linear, and ii) the fact that background lenses are themselves lensed by foreground lenses, thus changing the overall trajectory of light rays. These terms sometimes known as post-Born corrections will tend to Gaussianise the lensing fields since they characterise the introduction of random deflections along the light path which will in turn tend to diminish the impact of the non-linear clustering of matter. The mental image one could form is that of clustered chunks of matter blurred by these lensing terms. Those corrections were shown to be of little importance in the case of the PDF of the convergence field at scales relevant for perturbation theory ($\theta \sim 10$ arcmin) and relatively small redshifts relevant for upcoming cosmic shear surveys such as Euclid/LSST ($z \sim 1 - 2$) (Barthelemy et al. 2020b).

Extending this result to the aperture mass PDF boils down to computing the post-Born corrections on the dominant non-Gaussian term that makes the PDF, namely its reduced skewness. Using equation (41) of Barthelemy et al. (2020b) the post-Born correction to the third moment of a filtered convergence field is given by

$$\langle M_{\text{ap}}^3 \rangle_{\text{corr}} = -\frac{12}{(2\pi)^4} \int \frac{dz'}{H(z')} \omega_{\text{null}}(z', z) \frac{dz''}{H(z'')} \omega_{\text{null}}(z'', z) \omega(z', z'') \left[ \int \frac{d^2 \ell_1 d^2 \ell_2}{[D(z')D(z'')]^{2}} P \left( \frac{\ell_1}{D(z')}, z' \right) P \left( \frac{\ell_2}{D(z''), z''} \right) H(\ell_1, \ell_2) W(\ell_1) W(\ell_2) W(\ell_1 + \ell_2) \right],$$ \hspace{1cm} \text{(F1)}

where

$$H(\ell_1, \ell_2) = \frac{\ell_1 \cdot \ell_2}{\ell_2^2} + \frac{(\ell_1 \cdot \ell_2)^2}{\ell_1^2 \ell_2^2},$$ \hspace{1cm} \text{(F2)}

and

$$W(\ell_1) = 2 J_1(\theta \ell_1) / \theta_1 - 2 J_1(\theta \ell_2) / \theta_2,$$ \hspace{1cm} \text{(F3)}

is the $M_{\text{ap}}$ window function in Fourier space and $J_1$ is the first order Bessel function of the first kind. The subscript (null) indicates where to input the nulled lensing kernel instead of the usual one if one implements nulling. Equation (F1) is general enough so that any window function applied on the convergence field can be used to get the post-Born corrective term to the skewness. Furthermore when involving top-hat windows, specific properties of Bessel functions can be used to further simplify the expression

$$\int_0^{2\pi} d\varphi W_{TH}(\ell_1 + \ell_2) \left[ 1 - \cos^2(\varphi) \right] = \pi W_{TH}(\ell_1) W_{TH}(\ell_2)$$ \hspace{1cm} \text{(F4)}
Thus plugging equations (F4) and (F5) yields

\[
\begin{aligned}
\langle M_{\text{ap}}^g \rangle_{\text{corr}} &= -\frac{12}{\pi^2 \theta_1^2 \theta_2^2} \int_0^\pi \frac{d\varphi}{H(z')} \omega_{\text{null}}(z', z)^2 \int_0^\pi \frac{dz''}{H(z'')} \omega_{\text{null}}(z'', z) \\
&\quad \times P \left( \frac{\ell_1}{D(z')} - z'' \right) \left\{ \langle \theta_1 J_1(\ell_1 \theta_2) - \theta_2 J_1(\ell_2 \theta_1) \rangle \theta_1 J_1(\ell_2 \theta_2) \\
&\quad - \theta_2 J_1(\ell_2 \theta_1) \rangle \theta_2 J_2(\ell_1 \theta_1) - J_0(\ell_1 \theta_1) \rangle J_1(\ell_2 \theta_2) + \theta_1 J_0(\ell_1 \theta_2) \rangle J_2(\ell_1 \theta_2) \}
\end{aligned}
\]

For a difference of top-hats with opening angles \( \theta_1 \) and \( \theta_2 = 2\theta_1 \), we obtain the correction on the nulled \( M_{\text{ap}} \) skewness due to lens-lens coupling and geodesic deviation shown in Fig. F1 for source redshifts located at \( z_s = 1.2, 1.4 \) & 1.6 as a function of \( \theta_1 \). As expected, the correction is shown to i) Gaussianise the field and ii) be very small – sub-percent – which is not surprising since reducing the lensing kernel diminishes the importance of couplings between lenses.

**F2 Reduced shear correction**

Assuming that the intrinsic ellipticity of galaxies has no preferred orientation, note that this is not the case in the presence of intrinsic galaxy alignments, the observed ellipticity is an unbiased estimator of the reduced shear \( g = \gamma / (1 - \kappa) \) rather than the shear itself. This implies that rather than equation (6), the observed physical aperture mass is given by

\[
M_{\text{ap}}^g(\theta) = \int d^2 \vartheta' Q_\theta \left( \vartheta' \right) g_\theta \left( \theta - \vartheta' \right)
\approx M_{\text{ap}}(\theta) + \int d^2 \vartheta' Q_\theta \left( \vartheta' \right) \gamma_\theta \left( \theta - \vartheta' \right) \kappa \left( \theta - \vartheta' \right)
\approx M_{\text{ap}}(\theta) + \delta M_{\text{ap}}(\theta).
\]

We derive in this section the leading correction to the aperture mass skewness that accounts for the reduced shear. This is merely a perturbation theory calculation which still somewhat relies on the fact that the convergence \( \kappa \) is small and described by linear perturbation theory. Though this is rather straightforward and very useful for an estimation of the amplitude of the effect, this is not at the same level that what can be performed using large deviation theory to account for the effect as is shown without any projection effects in Reimberg & Bernardeau (2018).

First let us recall that one can deduce from their definition as derivatives of the projected gravitational potential that the convergence and the shear are equal in harmonic space up to a phase. Denoting \( f \) quantities in harmonic space (\( f \) being either the shear or the convergence field in our case) and introducing the wave vector \( \ell = (\ell \cos \varphi, \ell \sin \varphi) \) conjugate of \( \vartheta = (\vartheta \cos \varphi, \vartheta \sin \varphi) \) we have

\[
\gamma(\ell) = \tilde{\kappa}(\ell)e^{2i\varphi}. \tag{F9}
\]

Since the tangential shear is defined as

\[
\gamma_\theta(\vartheta) = -\Re \left( \gamma(\vartheta)e^{-2i\varphi} \right), \tag{F10}
\]

then it can be expressed as a function of the convergence field through

\[
\gamma_\theta(\vartheta) = -\int \frac{d^2 \ell}{(2\pi)^2} \cos(2(\vartheta - \varphi)) \cos(2\varphi) \cdot \delta \tilde{\kappa}(\ell). \tag{F11}
\]

Up to leading order, the skewness of the physical aperture mass is then

\[
\langle (M_{\text{ap}}^g)^3 \rangle \approx \langle M_{\text{ap}}^3 \rangle + 3 \langle M_{\text{ap}}^2 \delta M_{\text{ap}} \rangle \tag{F12}
\]

where the reduced shear correction is thus written as

\[
\langle M_{\text{ap}}^2 \delta M_{\text{ap}} \rangle = -\int d^2 \vartheta_U \vartheta_2 \langle \vartheta_1 \rangle \int \frac{d^2 \ell_2}{(2\pi)^2} \delta \ell_2 \\
\int \frac{d^2 \ell_1}{(2\pi)^2} \delta \ell_1 \cdot \vartheta_2 \cdot \vartheta_1 \int d^2 \vartheta Q_{\theta_2} \langle \vartheta_1 \rangle \int \frac{d^2 \ell'}{(2\pi)^2} \delta \ell' \cdot \vartheta \cos(2(\vartheta_2 - \varphi)) \langle \tilde{\kappa}(\ell_1) \tilde{\kappa}(\ell_2) \tilde{\kappa}(\ell') \rangle. \tag{F13}
\]

Again stopping at leading order, we only consider the linear evolution of the matter density fluctuation which means that the \( \kappa \) field is Gaussian. This allows to use Wick’s theorem to compute the correlator in equation (F13) which becomes

\[
\langle \tilde{\kappa}(\ell_1) \tilde{\kappa}(\ell_2) \tilde{\kappa}(\ell') \rangle \langle \tilde{\kappa}(\ell') \rangle = C_{\ell_1}^\kappa(\ell_1) C_{\ell_2}^\kappa(\ell_2) \delta D(\ell_1 + \ell_2) \delta D(\ell + \ell') + C_{\ell_1}^\kappa(\ell_1) C_{\ell_2}^\kappa(\ell_2) \delta D(\ell_1 + \ell) \delta D(\ell_2) \delta D(\ell_2 + \ell') + C_{\ell_1}^\kappa(\ell_1) C_{\ell_2}^\kappa(\ell_2) \delta D(\ell_1 + \ell + \ell') \delta D(\ell_2 + \ell_2). \tag{F14}
\]

The first term will yield zero because of the integration of

\[
\text{MNRAS 000, 1–20 (2020)}
\]
the cosine and the 2 other terms yield the same contribution. Carrying out the integration other \( \ell_1 \) and \( \ell_2 \) is straightforward thanks to the presence of the Dirac delta functions and we arrive at

\[
\langle M_{\text{ap}}^2 \delta M_{\text{ap}} \rangle = -2 \int d^2 \vartheta_1 U_\vartheta(\vartheta_1) \int d^2 \vartheta_2 U_\vartheta(\vartheta_2) \int d^2 \vartheta Q_\vartheta(\vartheta) \frac{d^2 \ell'}{(2\pi)^2} e^{i \ell' \cdot (\vartheta - \vartheta_1)} C^s_\ell(\ell) \cos(2(\varphi - \varphi_s)) \int d^2 \ell'' \frac{d^2 \ell''}{(2\pi)^2} e^{i \ell'' \cdot (\vartheta - \vartheta_2)} C^s_{\ell''}(\ell'') \cos(2(\varphi - \varphi_s)). \tag{F15}
\]

The integration over \( \vartheta_1, \vartheta_2 \) leads to the Fourier expression of the aperture mass (convergence-) filter and all that remains is to perform the angular integrations over \( \vartheta, \ell \) and \( \ell' \). These yield expressions that correspond to the integral definition of the Bessel functions of the first kind

\[
J_n(|x|) = \frac{1}{2\pi(\ell)^n} \int_{-\pi}^{\pi} d\varphi e^{i(\ell \cos(\varphi) - n\varphi)}, \quad n \in \mathbb{N}, \ x \in \mathbb{R}. \tag{F16}
\]

The expression for the reduced shear correction to the skewness of the aperture mass is thus

\[
\langle (M_{\text{ap}}^s)^3 \rangle_{\text{corr}} = \frac{3}{\pi} \int d\ell' C^s_\ell(\ell)W(\ell) \int d\ell'' C^s_{\ell''}(\ell')W(\ell') \int d\vartheta Q_\vartheta(\vartheta)J_0(\vartheta\ell')J_2(\vartheta\ell). \tag{F17}
\]

We now evaluate this correction in the same configuration as the post-born correction, that is nulling with evolving opening angles and plot the resulting correction in Fig. F1. Finally, note that in the case the nulling, the procedure is only exact for the convergence and shear fields and thus another term accounting for this issue should also be present when introducing the reduced shear correction. Taking into account all corrections that arise because of inaccuracies of the procedure in realistic settings is left for future work.

### F3 Magnification bias correction

Individual galaxies can be (de)magnified and thus their flux is (de)increased. At the flux limit of a survey, this can cause fainter sources to be included in the observed sample while they would, in the absence of lensing, be excluded. At the same time, the density of galaxies in the small region around this source appears reduced (increased) since it is also (de)magnified. As such, the net effect depends on the slope of the intrinsic, unlensed, galaxy luminosity function at the survey’s flux limit. This is known as the magnification bias. We follow the prescription given in Deshpande et al. (2020) for the resulting *observed* shear (and not the reduced shear since we are looking at the leading order correction) which reads

\[
\gamma_{\text{obs}} = \gamma + \gamma \delta^s + (5s - 2)\gamma \kappa, \tag{F18}
\]

where \( \delta^s \) is the intrinsic, unlensed, galaxy overdensity at the source (or in the redshift bin),

\[
s = \left. \frac{\partial \log_{10} n(z_s, m)}{\partial m} \right|_{m_{\text{lim}}}, \tag{F19}
\]

\[n(z_s, m)\text{ is the true distribution of galaxies, evaluated at the source (central redshift of the bin) and at a given magnitude (luminous flux) } m, \text{ and } m_{\text{lim}} \text{ is the survey’s limiting magnitude. The leading order correction to the skewness coming from this effect is exactly the same as in the case of the reduced shear in equation (F17) up to a factor } (5s - 2). \text{ Indeed the } \gamma \delta^s \text{ term will yield a zero contribution in any correlator if the lenses and sources do not overlap. This is the case when considering source planes or very narrow redshift bins but more importantly this is also the case when applying the BNT transform to any tomographic bins (Bernardeau et al. 2014c).} \]