Large deviations theory approach to cosmic shear calculations: the one-point aperture mass

Paulo Reimberg, Francis Bernardeau

Large deviations theory approach to cosmic shear calculations: the one-point aperture mass

Paulo Reimberg*

Sorbonne Universités, UPMC Univ Paris 6 et CNRS, UMR 7095,
Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France

Francis Bernardeau

Sorbonne Universités, UPMC Univ Paris 6 et CNRS, UMR 7095,
Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France and
CEA - CNRS, UMR 3681, Institut de Physique Théorique, F-91191 Gif-sur-Yvette, France

This paper presents a general formalism that allows the derivation of the cumulant generating function and one-point Probability Distribution Function (PDF) of the aperture mass ($\hat{M}_{ap}$), a common observable for cosmic shear observations. Our formalism is based on the Large Deviation Principle (LDP) applied, in such cosmological context, to an arbitrary set of densities in concentric cells. We show here that the LDP can indeed be used for a much larger family of observables than previously envisioned, such as those built from continuous and nonlinear functionals of the density profiles. The general expression of the observable aperture mass depends on reduced shear profile making it a rather involved function of the projected density field. Because of this difficulty, an approximation that is commonly employed consists in replacing the reduced shear by the shear in such a construction neglecting therefore non-linear effects. We were precisely able to quantify how this approximation affects the $\hat{M}_{ap}$ statistical properties. In particular we derive the corrective term for the skewness of the $\hat{M}_{ap}$ and reconstruct its one-point PDF.

I. INTRODUCTION

Gravitational lensing has been shown to be a very efficient way of exploring the properties of the mass distribution at large scale. It indeed provides information about the gravitational potential that light rays go through, from sources to observer. Although it is fair to say that the most spectacular consequences of such phenomena are the strong lensing effects, with the occurrences of multiple images and large arc-like deformation of images of background objects, the most fruitful regime in the context of cosmological observation is the one of the cosmic shear: weak lensing effects are indeed ubiquitous and are induced by the large-scale structure of the universe as a whole. The first evidences that large-scale structure can coherently affect the shape of background galaxies have been presented in year 2000 in a series of compelling results, [1–3]. The evidence is based on the fact that such deformations are expected to obey a specific geometrical property, namely the absence of parity-odd contributions (i.e. negligible B-modes). These results opened the way to a systematic use of such observations to map the large-scale structure of the universe and explore its statistical properties. To be more specific, in such a regime (in the absence of critical region), the information provided by cosmic shear observations are encoded in the elements of a deformation matrix, the convergence and shear fields, that describe the magnification and deformation of the shape of light beams. The reconstruction of

such maps from background galaxy shapes would provide in principle direct information about the projected mass [4–6]. This is key to a large part of the core programs of projects such as the CFHTLenS[1], the Dark Energy Survey (DES)[2], LSST[3], and Euclid[4].

Cosmic shear observations are usually exploited with the help of the shear two-point correlation function (or equivalently its power spectrum). This is what the approach which is usually adopted such as in [7] for the CFHTLenS survey and in [8] for the DES survey. There exist however alternative approaches that give complementary information. This is in particular the case of the convergence one-point Probability Distribution Function (PDF) or rather the PDF of the aperture mass (which is a specific filtering of the convergence map we define below) and its first few moments as exploited in [9] for the CFHTLS or in [10] for the DES survey.

This paper is in this line of investigation. It does not aim at quantifying the efficiency of such measurements in constraining the cosmological models but at showing the one-point PDF of the aperture mass can be computed from first principles in a given cosmological context. Such a computation makes use of the Large Deviations Principle (LDP), that is the central definition on the theory of large deviations, and that we will describe more specifically in the following. In broad terms, the theory of large deviations [11–13] is a branch of probability theory that

---

*Electronic address: reimberg@iap.fr

deals with rate at which probabilities of certain events decay as a natural parameter of the problem varies [14], and is applied in variety of domains in mathematics and theoretical physics, specially in statistical physics both for equilibrium and non-equilibrium systems (see for instance [15] for a review paper on the subject, or App. B for a short introduction). The application of the LDP to Large Scale Structure cosmology has been formalized in the last few years [16–19] and will be employed here in the context of cosmic shear observations. This is not the first time that arguments related to the theory of large deviations is used in this context, and we highlight [20–23] in particular.

In this paper we focus on aperture mass statistics and extend the results that had been obtained in two ways: the aperture mass is defined with the help of more realistic filters, and we show that one can take into account the fact that only the reduced shear can be measured, not the shear itself. Cosmic shear observations are indeed based on the measurements of background galaxy shapes, more specifically on the amplitude and direction of their deformation, and what we have access to are ratios of the deformation matrix elements, i.e; the reduced shear which is then the ratio (shear)/(1-convergence). Moreover, from such observation one can easily build the Laplacian of the associated (reduced) convergence field. But it is not possible to unambiguously recover the convergence field itself. For circular symmetric filters, the convergence can only be recovered in compensated filters (which filter out constant fields and fields with constant gradients) which can be viewed as projected mass maps in aperture[5], that is *aperture mass* (Map) [24, 25].

In general observations should then be viewed as non local, non linear transformation of the convergence field. In the literature the reduced shear is usually replaced by the shear itself arguing that the convergence ought to be generally small for cosmic shear observations. It dramatically simplifies the problem as the observed Map is made a linear transformation of the projected mass field. This is however an a priori unjustified simplification in the context of the theory of large deviations. It can be shown that the statistical properties of quantities that respect certain symmetries (which here will be the circular symmetry) can be computed in the small variance limit even though some events strongly depart form the variance. What the LDP can account for is precisely the impact of excursions of large values of the convergence.

The general techniques for computing one-point PDF of densities filtered with top-hat filters has been developed in a long series of papers [17, 26–28]. The derivation of exact results for continuous filters escaped however this formalism[6]. The central ingredient of the developments presented here is the construction of general symmetric filtering of the density field as the continuous limit of a weighted composition of concentric top-hat filters, as introduced in [18]. The second main point of this paper is to show that the LDP can be applied to a non-linear functional of the density profile and namely to the one-point PDF of the observed Map. Again this construction will make use of the general formulation developed in [18]. Our work therefore extends previous results in two fundamental ways.

The application of the LDP is based on a number of ingredients we will detail in section II. The line of reasoning we follow (similar to [20]) is based on the fact that cosmic shear observations are akin to observation in long cylindrical cells. In the following we will simply assume that the fluctuations along the radial direction have simply been integrated out. The LDP is then based on the assumption that the leading configurations in the initial field that lead, after nonlinear evolution, to a given circular constraint obey the same circular symmetry. Under this assumption (and before shell crossing) it is then possible to map the initial configuration to the final one. The setting of the LDP is then based on the following ingredients:

- One should first define the rate function of the variables that define the initial field configuration. We will assume here Gaussian initial condition to define their covariance matrix.

- One should then specify the mapping between the initial field configuration (mass profile) and the final mass profile. It will make use of the 2D cylindrical collapse (or rather one approximation of it).

- We are then in position to write the observable – say the Map defined with a specific filter – as a functional form of the final and therefore initial mass profile.

The rate function of the Map variable is then obtained through a minimization procedure we detail in section III. The results are discussed in section IV and in particular we comment on the impact of the differences between Map and observed Map statistics at the level of the rate function, the cumulant generating function and the resulting one-point PDF.

II. THE LDP APPLIED TO THE CONVERGENCE FIELD

The aim of this section is to introduce the necessary ingredients required for the implementation of the LDP. We first review the nature of the observable we are interested in.

---

[5] Conversely it can be shown that such Map fields can be written in terms of integrals of well chosen components the reduced shear field.

[6] The derivation of the skewness when gaussian filtering is employed, for example, was a remarkable tour de force [29].

A. The convergence

Lensing effects are generally classified in strong and weak. On the first case arcs and multiple images are observable as consequence of caustics and cusps on the observers past-light cone due to gravitation. Weak gravitational lensing is associated to smooth deformations on the observers past-light cone and the effects are less dramatic. Gravitational lensing on distant galaxies can be observed through the distortion of shape and size of the light sources parametrized by the *shear* γ and *convergence* κ, respectively. For scalar, linear perturbations of the FLRW spacetime, convergence and shear can be expressed as derivatives of a *lens potential* (see App. A). The convergence, in particular, is given by $2\kappa = (\nabla_1\nabla_1 + \nabla_2\nabla_2)\phi_L$, where $\nabla_{1,2}$ are derivatives on a spacelike plan perpendicular to the direction of propagation of the light beam. If we assume Born approximation, neglect lens-lens couplings and extrapolate the 2D Laplacian that appears on the definition of κ to the 3D Laplacian, we can use the Poisson's equation to link convergence and matter density fluctuations as (see for instance [5]):

$$\kappa(\boldsymbol{\theta}) = \int_0^{\chi_S} \mathrm{d}\chi_L \, w(\chi_S, \chi_L) \delta(\mathcal{D}_0(\chi_L)\boldsymbol{\theta}, \chi_L). \quad (1)$$

Here χ_S is the comoving distance to the sources, and

$$w(\chi_S, \chi_L) := \frac{3}{2}\Omega_m \frac{H_0^2}{c^2} \frac{\mathcal{D}_0(\chi_S - \chi_L)\mathcal{D}_0(\chi_L)}{a(\chi_L)\mathcal{D}_0(\chi_S)} \quad (2)$$

with $\mathcal{D}_0$ is defined in Eq. (A3). We observe that $w(\chi_S, \chi_L)$ is a positive function, and therefore by the integral Mean-Value theorem, there exists χ_L^*, $0 < \chi_L^* < \chi_S$, such that

$$\kappa(\boldsymbol{\theta}) = w_{\mathrm{eff}}(\chi_S) \, \delta_{2D}(\mathcal{D}_0(\chi_S)\boldsymbol{\theta}, \chi_S) \quad (3)$$

we will not keep the dependence on χ_S explicitly in $\kappa(\boldsymbol{\theta})$ to make the notation shorter. We define $w_{\mathrm{eff}}(\chi_S)$ explicitly as:

$$w_{\mathrm{eff}}(\chi_S) := \frac{3}{2}\Omega_m \frac{H_0^2}{c^2} \frac{\mathcal{D}_0(\chi_S - \chi_L^*)\mathcal{D}_0(\chi_L^*)}{a(\chi_L^*)\mathcal{D}_0(\chi_S)}. \quad (4)$$

This corresponds effectively to existence os a *lens plan* at χ_L^* where all the lensing mass would be concentrated. We could have defined a *projected density contrast* instead, as in [20], but the assumption of a lens plan does not weaken our present argument. In Einstein-de Sitter universe, the possible values of $w_{\mathrm{eff}}(z_S)$ are shown in Fig. 1 as function of the possible z_L^*s.

The quantity $\delta_{2D}(\mathcal{D}_0(\chi_S)\boldsymbol{\theta}, \chi_S)$ is given by:

$$\delta_{2D}(\mathcal{D}_0(\chi_S)\boldsymbol{\theta}, \chi_S) = \int_0^{\chi_S} \mathrm{d}\chi' \delta(\mathcal{D}_0(\chi')\boldsymbol{\theta}, \chi'), \quad (5)$$

and plays the role of projected density contrast. We will also assume a small angle approximation and consider
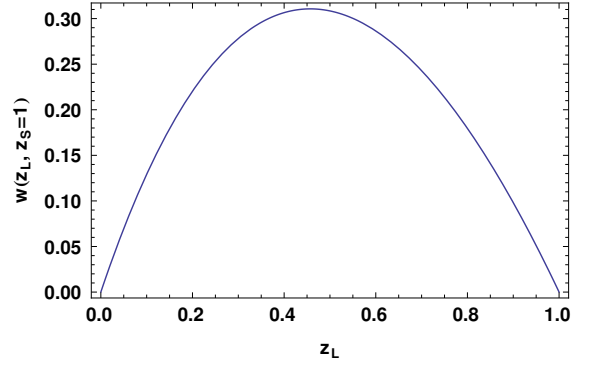

FIG. 1: Projection factor $w(z_L, z_S)$ defined in Eq. (2) in Einstein- de Sitter universe for $z_S = 1$. The effective value w_{eff} is the particular values of $w(z_L, z_S)$ at $z_L = z_L^*$ corresponding to the redshift of the effective lens plan.

that the integral on Eq. (5) is performed in a cylindrical region instead of a conic one. We refer to [20, 28, 30] to a detailed discution of the approximations assumed here.

For reasons that will be explicit in the next paragraph we are led to consider the smoothed convergence over cylindrical regions with aperture scales θ_i:

$$\kappa_<(\theta_i) := \int_0^{\theta_i} \frac{\mathrm{d}^2\boldsymbol{\theta}}{\pi\theta_i^2}\kappa(\boldsymbol{\theta}) \quad (6)$$

for a given sequence of scales $\theta_1, \theta_2, \ldots, \theta_N$ of particular interest. We observe that, except for a multiplicative factor, this corresponds to the smoothing of the surface density δ_{2D}. We shall, then, define

$$\hat{\kappa}_<(\theta) = \frac{\kappa_<(\theta)}{w_{\mathrm{eff}}(\chi_S)} \quad (7)$$

as a normalized convergence.

B. The rate function of the initial field configuration

We will consider the evolution of the effective 2D density field on the lens plane: as a first step, a Gaussian distributed density field τ_{2D} is set, and for this field we define $\hat{\kappa}_<^{lin}(\theta)$ as in Eq. (7), but using τ_{2D} in Eq. (3). We want to compute the elements $\Sigma_{ij} := \langle \hat{\kappa}_<^{lin}(\theta_i)\hat{\kappa}_<^{lin}(\theta_j)\rangle$, $1 \le i, j \le N$ of the covariance matrix Σ for this gaussian field. On small-angle approximation one obtains [28]:

$$\Sigma_{ij} = \int_0^\infty \frac{\mathrm{d}k_\perp k_\perp}{2\pi}P(k_\perp)W(\mathcal{D}_0\theta_i k_\perp)W(\mathcal{D}_0\theta_j k_\perp) \quad (8)$$

where

$$W(x) = \frac{2J_1(x)}{x} \quad (9)$$

is the Fourier transform of the top-hat filter in two dimensions. We normalize D_+ to be unity at current time. We assume here $P(k_\perp) = Ak_\perp^n$, $-2 \le n \le 1$.

As observed in appendix B, the collection $\{\hat{\kappa}_<^{lin}(\theta_i)\}_{1\leq i\leq N}$ of correlated gaussian random variables obeys the LDP with rate function:

$$I(\hat{\kappa}_<^{lin}(\theta_1),\ldots,\hat{\kappa}_<^{lin}(\theta_N)) = \frac{\sigma^2(\theta_N)}{2}\sum_{ij}\Xi_{ij}\,\hat{\kappa}_<^{lin}(\theta_i)\,\hat{\kappa}_<^{lin}(\theta_j) \tag{10}$$

where $\Xi = \Sigma^{-1}$, and $\sigma^2(\theta_N) = \Sigma_{NN}$. When we take the limit $\sigma^2(\theta_N) \to 0$, the rate function determines the exponential decay rate for the probability density function associated to the random variables.

C. The mapping between the initial configuration and the final configuration

We assume now that the τ_{2D} is a density fluctuation produced by a gas of non-interacting particles obeying continuity, Euler and Poisson equations with azimuthal symmetry, i.e., a cylindrical collapse. If this dynamics can be solved, a map connecting linear and non-linear overdensities can be established. If we consider matter contained in a cylindrical region, Gauss theorem will provide us the relation of the radii given initial and final density. For what concerns this work, it is sufficient to know that the normalized non-linear density can be approximated in terms of the linear density as [28]:

$$\zeta(\tau_{2D}) = \frac{1}{\left(1 - \frac{\tau_{2D}}{\nu}\right)^\nu}, \qquad \nu = \frac{\sqrt{13}-1}{2}. \tag{11}$$

We can therefore construct a new family of random variables to describe the convergence produced by the non-linear evolution of τ_{2D}:

$$\hat{\kappa}_<(\vartheta_i) = \zeta(\hat{\kappa}_<^{lin}(\theta_i)) - 1 \qquad \vartheta_i = \theta_i/\zeta(\hat{\kappa}_<^{lin}(\theta_i))^{1/2}. \tag{12}$$

Since we assume no shell-crossings, the angular scales ϑ_i are related to the initial scales θ_i by the constraint of mass conservation inside a given shell. The family of random variables $\{\hat{\kappa}_<(\vartheta_i)\}_{1\leq i\leq N}$ is obtained from $\{\hat{\kappa}_<^{lin}(\vartheta_i)\}_{1\leq i\leq N}$ by the continuous function ζ, and therefore new family of random variables also obeys the LDP as a consequence of the *contraction principle* in Large Deviations Theory (see appendix B).

D. The rate function of the final field configuration

The contraction principle states that the rate function for the new family will be given by (see Eq. (B4)):

$$\Psi(\hat{\kappa}_<(\vartheta_1),\ldots,\hat{\kappa}_<(\vartheta_N)) = \inf_{\hat{\kappa}_<^{lin}} I(\hat{\kappa}_<^{lin}(\theta_1),\ldots,\hat{\kappa}_<^{lin}(\theta_N)), \tag{13}$$

where $\inf_{\hat{\kappa}_<^{lin}}$ stands for the infimum taken over the collection $\{\hat{\kappa}_<^{lin}(\theta_i)\}_{(1\leq i\leq N)}$ such that $\hat{\kappa}_<(\vartheta_i) = \zeta(\hat{\kappa}_<^{lin}(\theta_i))-1$.

In the domain in which ζ is bounded we can perform the inversion $\hat{\kappa}_<^{lin}(\theta_i) = \zeta^{-1}[1 + \hat{\kappa}_<(\vartheta_i)]$, that we may also write as $\hat{\kappa}_<^{lin}(\hat{\kappa}_<(\vartheta_i))$. We can therefore write,

$$\Psi(\hat{\kappa}_<(\vartheta_1),\ \ldots\ ,\hat{\kappa}_<(\vartheta_N)) = \frac{\sigma^2(\vartheta_N)}{2}$$
$$\times \sum_{ij}\Xi_{ij}\hat{\kappa}_<^{lin}(\hat{\kappa}_<(\vartheta_i))\hat{\kappa}_<^{lin}(\hat{\kappa}_<(\vartheta_j)). \tag{14}$$

Again $\Xi = \Sigma^{-1}$, Σ being the matrix whose elements are given in Eq. (8).

The Legendre-Fenchel transform of the rate function is the *scaled cumulant generating function* (SCGF), from which all the cumulants can, in principle, be derived (see appendix B).

E. The single cell case

In order to summarize and illustrate the rate function, SCGF, their relations and role on the derivation of observable quantities, we will consider the convergence filtered at one given scale, i.e., we take $N = 1$ in (14). The rate function in this case will be given by

$$\Psi(\hat{\kappa}_<(\vartheta)) = \frac{\sigma^2(\vartheta)\,(\hat{\kappa}_<^{lin}(\hat{\kappa}_<(\vartheta)))^2}{2\sigma^2(\vartheta\,\zeta^{1/2}(\hat{\kappa}_<^{lin}(\theta)))}. \tag{15}$$

If $P(k) \propto k^n$, then $\sigma^2(x) \propto x^{-(n+2)}$ in 2D dynamics. We observe from the graph of the function $\Psi(\hat{\kappa}_<)$ shown in Fig. 2 that this function is not globally convex. Indeed there is a critical value $\hat{\kappa}_<^c$ where there is a change of convexity. For $-2 \leq n \leq 1$, however, $\hat{\kappa}_<^c > 0$ indicating that the rate function is convex in a neighborhood of the origin.
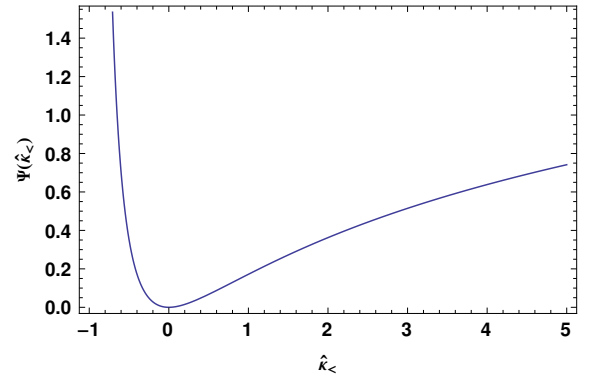

FIG. 2: The rate function given in Eq. (15) for $n = -1.5$. The rate function is convex around the origin but changes the convexity for $\hat{\kappa}_< \approx 0.8$.

When $\sigma^2 \to 0$, the scaled cumulant generating function is the Legendre-Fenchel transform of the rate function, i.e,

$$\varphi(\lambda) = \sup_{\hat{\kappa}_<}[\lambda\,\hat{\kappa}_< - \Psi(\hat{\kappa}_<)]. \tag{16}$$

The quantities φ and Ψ are said to be *convex conjugate*, and Eq. (16) is written simply as $\varphi = \Psi^*$ in some references. If Ψ is globally convex, then $\Psi = \varphi^* = \Psi^{**}$ (i.e., convex conjugation is involutive on the space of convex functions). If Ψ is not globally convex, then Ψ^{**} produces only the convex envelope of Φ. Moreover, $\varphi = \Psi^*$ has points of non-differentiability when Ψ loses convexity. Our rate function, as observed, is not globally convex for all values of the spectral index n, what requires careful analysis on the inversion of φ to obtain the PDF [17, 19].

On the strictly convex domain of Ψ, however, the Legendre-Fenchel transform reduces to the classical Legendre transform and therefore $\lambda = \frac{\partial \Psi(\hat{\kappa}_<)}{\partial \hat{\kappa}_<}$. If the scaled cumulant generating function is C^∞ around the origin, then all the (scaled) cumulants can be obtained from the Taylor expansion of $\varphi(\lambda)$:

$$\varphi(\lambda) = \sum_{p=1}^{\infty} \hat{s}_p \frac{\lambda^p}{p!} \,. \tag{17}$$

Here $\hat{s}_p = \langle \hat{\kappa}_<^p \rangle_c / \langle \hat{\kappa}_<^2 \rangle_c^{p-1}$. Since p-th cumulant is an homogeneous functions of degree p, $s_p := \langle \kappa^p \rangle_c / \langle \kappa^2 \rangle_c^{p-1} = w_{\text{eff}}(\chi_S)^{2-p} \hat{s}_p$. For the skewness ($p = 3$) we have:

$$s_3(\vartheta) = \frac{1}{w_{\text{eff}}(\chi_S)} \left[\frac{36}{7} - \frac{3}{2}(n+2) \right] \tag{18}$$

from where we can depict the inversely proportional dependence of the skewness on Ω_{m}. As it was shown in [30], $s_3 \approx 42 \, \Omega_{\text{m}}^{-0.8}$ for $z_S \approx 1$.

III. THE LDP FORMULATION APPLIED TO THE OBSERVED APERTURE MASS

We have considered until now a family of random variables produced by filtering the convergence field with a family of top-hat filters associated to scales $\mathcal{D}_0 \vartheta_1, \ldots, \mathcal{D}_0 \vartheta_N$ for a given lensing configuration. More general filtering schemes can be proposed, and a very convenient example is the so called *aperture mass*, which is produced by the convolution of the convergence field with a compensated filter $U(x)$, i.e., a filter with the property $\int \mathrm{d}x \, x \, U(x) = 0$. In order to keep working with normalized quantities, we shall define the *normalized aperture mass* as in [20]:

$$\hat{M}_{ap} = U * \hat{\kappa} \,. \tag{19}$$

The scale on which the convergence field is filtered here is determined by a parameter on the definition of $U(x)$. We will compute the aperture mass at the origin here.

A. Filters for convergence and shear

We should first remark that the aperture mass is defined in terms of a filtering of the convergence field. We

would like, instead, to use the top-hat filtered quantity $\hat{\kappa}_<$ on the definition. We remark that the differentiation of Eq. (6) yields $\hat{\kappa}(\vartheta) = \kappa(\vartheta)/w = \hat{\kappa}_<(\vartheta) + \frac{\vartheta}{2} \hat{\kappa}'_<(\vartheta)$. We can thus use integration by parts to write

$$\hat{M}_{ap} = \int \mathrm{d}\vartheta \, \vartheta \, U(\vartheta) \, \hat{\kappa}(\vartheta) = \int \mathrm{d}\vartheta \, \tilde{U}(\vartheta) \, \hat{\kappa}_<(\vartheta) \tag{20}$$

where $\tilde{U}(\vartheta) = -\frac{\vartheta^2}{2} U'(\vartheta)$. The filter $\tilde{U}$ satisfies $\int \mathrm{d}\vartheta \, \tilde{U}(\vartheta) = 0$. Going one step further, we can express the $\hat{M}_{ap}$ in terms an specific filter acting on the shear [25]. For axially symmetrical lenses, we can relate $\hat{\kappa}$, $\hat{\kappa}_<$, and $\hat{\gamma} \, (= \gamma/w)$ as:

$$\hat{\gamma}(\vartheta) = \hat{\kappa}_<(\vartheta) - \hat{\kappa}(\vartheta) = -\frac{\vartheta}{2} \hat{\kappa}'_<(\vartheta) \,. \tag{21}$$

Assume that there exists a filter $\tilde{Q}$ such that

$$\begin{aligned} \hat{M}_{ap} &= \int \mathrm{d}\vartheta \, \tilde{Q}(\vartheta) \, \hat{\gamma}(\vartheta) \\ &= \int \mathrm{d}\vartheta \, \frac{\tilde{Q}(\vartheta) + \vartheta \, \tilde{Q}'(\vartheta)}{2} \, \hat{\kappa}_<(\vartheta) \,, \end{aligned} \tag{22}$$

where the second line was obtained by integration by parts. The relation between the filters $\tilde{U}$ and $\tilde{Q}$ is, therefore,

$$\tilde{U}(\vartheta) = \frac{1}{2} \left(\tilde{Q}(\vartheta) + \vartheta \, \tilde{Q}'(\vartheta) \right) \,. \tag{23}$$

As before, $\tilde{Q}(\vartheta) = -\frac{\vartheta^2}{2} Q'(\vartheta)$. For analytical convenience, we will focus on the filter $Q(\vartheta) = \mathrm{e}^{-\vartheta^2/2}$. Other possible compensated filters can be found in the literature [20, 25, 31].

B. Conveniences and limitations of the aperture mass

The convergence field can only be obtained through inversion problem and is non-local. Shear field is more directly observed. As discussed in appendix A 1, observation of ellipticity fields yields a measure of the reduced shear ($g = \gamma/(1 - \kappa)$), from which the shear is usually obtained by assuming $\kappa \ll 1$ for weak lensing [5, 6]. This leads us to define the *observed aperture mass* as:

$$\hat{M}_{ap}^g := \int \mathrm{d}\vartheta \, \hat{g}(\vartheta) \, \tilde{Q}(\vartheta) \approx \hat{M}_{ap} \tag{24}$$

where $\hat{M}_{ap}$ is given in Eq. (22). The (*theoretical*) aperture mass $\hat{M}_{ap}$ has therefore the good analytical property of being expressible in terms of the shear or convergence, but the observable quantity is $\hat{M}_{ap}^g$. The theoretical and observable aperture masses should agree as long as the convergence is small. Being defined as the filtering of a

random field, however, the convergence can exhibit statistical large fluctuations, and therefore the statement $\kappa \ll 1$ is not well defined on itself.

We seek, then, to quantify the role of large deviations on the reconstruction of statistical properties of the aperture mass.

C. General filter and the $\hat{M}_{ap}$ statistics

We have presented in Sec. II C the formalism based on the Large Deviations Theory to derive the SCGF associated to the top-hat filtering of the convergence field for a finite number of filtering scales. A general filtering can be approximated by a weighted composition of top-hat filters as introduced in [18], and we will develop this generalization to produce the SCGF for the $\hat{M}_{ap}$ and $\hat{M}_{ap}^g$.

We shall assume that the $\hat{M}_{ap}$ can be approximated by a finite collection of the top-hat filtered convergence $\hat{\kappa}_<(\vartheta_i)$ weighted by the series of coefficients $\tilde{U}_i$, $1 \le i \le N$ as:

$$\hat{M}_{ap} \approx \sum_{i=1}^N \tilde{U}_i(\vartheta_i)\, \hat{\kappa}_<(\vartheta_i). \qquad (25)$$

If this is the case, the $\hat{M}_{ap}$ is a linear combination of the $\hat{\kappa}_<(\vartheta_i)$'s and therefore, by the contraction principle, the rate function for the $\hat{M}_{ap}$ is given in terms of the initial rate function $I(\theta_1, \ldots, \theta_N)$ through a composition of continuous bounded functions. By the contraction principle (see Eq. (B4)),

$$\Psi(\hat{M}_{ap}) = \inf_{\hat{\kappa}_<^{lin}(\theta_i)} \left[I(\hat{\kappa}_<^{lin}(\theta_1), \ldots, \hat{\kappa}_<^{lin}(\theta_N)) \right.$$
$$\left. + \alpha \left(\hat{M}_{ap} - \sum_{i=1}^N \tilde{U}_i(\vartheta_i)\, \hat{\kappa}_<(\vartheta_i) \right) \right] \quad (26)$$

where α is a Lagrange multiplier and $\hat{\kappa}_<(\vartheta_i)$ is given in terms of $\hat{\kappa}_<^{lin}(\theta_i)$ by Eq. (12).

The weights $\tilde{U}_i$ are defined on the image of ζ, i.e., on the filtered non-linear field, but the inf has to be computed over the initial random variables $\hat{\kappa}_<^{lin}(\theta_1), \ldots, \hat{\kappa}_<^{lin}(\theta_N)$, both being related by the map ζ defined on Eqs. (11), (12).

We assume now that refinements in the partition defined by the filtering scales will eventually lead to the corresponding continuous limit,

$$\sum_{i=1}^N \tilde{U}_i(\vartheta_i)\, \hat{\kappa}_<(\vartheta_i) \to \int d\vartheta\, \tilde{U}(\vartheta)\, \zeta(\hat{\kappa}_<^{lin}(\theta))$$
$$= \int d\theta \frac{d\vartheta_{\hat{\kappa}_<^{lin}}}{d\theta} \tilde{U}[\vartheta_{\hat{\kappa}_<^{lin}}(\theta)]\, \zeta(\hat{\kappa}_<^{lin}(\theta)) \qquad (27)$$

where we stress the dependence of ϑ on $\hat{\kappa}_<^{lin}(\theta)$ by writing Eq. (12) as:

$$\vartheta_{\hat{\kappa}_<^{lin}}(\theta) = \theta\, \zeta^{-1/2}(\hat{\kappa}_<^{lin}(\theta)). \qquad (28)$$

D. The Scaled Cumulant Generating Functions

Our next goal is write the SCGF on the continuous limit. By the Varadhan's lemma given in Eq. (B5) the scaled cumulant generating function will be the continuous limit of

$$\varphi(\lambda) = \sup_{\hat{\kappa}_<^{lin}} \left[\lambda \sum_i^N \tilde{U}_i(\vartheta_i)\, \hat{\kappa}_<(\vartheta_i) - I(\hat{\kappa}_<^{lin}(\theta_1), \ldots, \hat{\kappa}_<^{lin}(\theta_N)) \right].$$
$$(29)$$

In order to write the continuous limit of the rate function $I(\hat{\kappa}_<^{lin}(\theta_1), \ldots, \hat{\kappa}_<^{lin}(\theta_N))$ given in Eq. (10), we have to give a continuous limit to the matrix Ξ. Let us assume, for this sake, the existence of an object $\xi(\theta', \theta'')$ such that

$$\int d\theta'\, \sigma^2(\theta, \theta')\, \xi(\theta', \theta'') = \delta_D(\theta - \theta'') \qquad (30)$$

where $\sigma^2(\theta, \theta')$ is given by Eq. (8) computed over continuous domains.

The desired continuous limit for the SCGF is, therefore,

$$\varphi(\lambda) = \sup_{\hat{\kappa}_<^{lin}} \left[\lambda \int d\theta\, \tilde{U}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta)) \frac{d\vartheta_{\hat{\kappa}_<^{lin}}}{d\theta}\, \zeta(\hat{\kappa}_<^{lin}(\theta)) \right.$$
$$\left. - \frac{\sigma_F^2}{2} \int d\theta\, d\theta'\, \hat{\kappa}_<^{lin}(\theta)\, \hat{\kappa}_<^{lin}(\theta')\, \xi(\theta, \theta') \right]. \quad (31)$$

Here $\sigma_F^2 = \int d\theta\, d\theta'\, \sigma^2(\theta, \theta')\, \tilde{U}(\theta)\, \tilde{U}(\theta')$.

As already observed, the normalized aperture mass $\hat{M}_{ap}$ can be obtained as in Eq. (20) through the convolution of $\hat{\kappa}_<$ and $\tilde{U}$, or as in Eq. (22) by convolving $\hat{\gamma}$ and $\tilde{Q}$, and the two different expressions are related by an integration by parts. Integrations by parts can also be applied on Eq. (31) to re-express it explicitly in terms of the convergence or shear with corresponding filters. Although equivalent, the different expressions of φ allows us perform distinct numerical implementations and check the performance of the solutions. A second need for a rewriting of Eq. (31) in terms of $\hat{\gamma}$ and $\tilde{Q}$ is to extend the SCGF for $\hat{M}_{ap}$ into the SCGF of $\hat{M}_{ap}^g$. We will call the different arrangement of variables *representations* of the SCGF.

1. SCGF on the convergence representation

The SCGF for $\hat{M}_{ap}$ in Eq. (31) is already given in terms of $\hat{\kappa}_<$ and $\tilde{U}$, but it can be expressed in a more

numerically suitable form. Let the filter $\tilde{V}(x)$ be defined by:

$$\tilde{V}(x) = \int_0^x dy \, \tilde{U}(y) \,. \tag{32}$$

It follows by integration by parts that:

$$\int_0^\infty d\theta \left(\tilde{U}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta)) \frac{d\vartheta_{\hat{\kappa}_<^{lin}}}{d\theta} \right) \zeta(\hat{\kappa}_<^{lin}(\theta))$$

$$= - \int_0^\infty d\theta \, \tilde{V}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta)) \frac{\partial \zeta(\hat{\kappa}_<^{lin}(\theta))}{\partial \hat{\kappa}_<^{lin}} \, \hat{\kappa}_<^{lin'}(\theta) \tag{33}$$

as long as the surface term $\tilde{V}(\vartheta_{\hat{\kappa}_<^{lin}}) \zeta(\hat{\kappa}_<^{lin}(\theta))|_0^\infty = 0$, what is the case if $U(x)$ is a compensated filter. The SCGF on the *convergence representation* is therefore:

$$\varphi^\kappa(\lambda) = - \inf_{\hat{\kappa}_<^{lin}} \left[\lambda \int d\theta \, \tilde{V}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta')) \frac{\partial \zeta(\hat{\kappa}_<^{lin}(\theta'))}{\partial \hat{\kappa}_<^{lin}} \, \hat{\kappa}_<^{lin'}(\theta') \right.$$

$$\left. + \frac{\sigma_F^2}{2} \int d\theta \, d\theta' \, \hat{\kappa}_<^{lin}(\theta) \, \hat{\kappa}_<^{lin}(\theta') \, \xi(\theta, \theta') \right]. \tag{34}$$

2. SCGF on the shear representation

As a direct consequence of Eq. (22) applied to Eq. (31), the SCGF on the *shear representation* is given by:

$$\varphi^\gamma(\lambda) = - \inf_{\hat{\kappa}_<^{lin}} \left[\lambda \int d\theta \, \tilde{Q}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta)) \frac{\vartheta_{\hat{\kappa}_<^{lin}}}{2} \frac{d\zeta(\hat{\kappa}_<^{lin}(\theta))}{d\theta} \right.$$

$$\left. + \frac{\sigma_F^2}{2} \int d\theta \, d\theta' \, \hat{\kappa}_<^{lin}(\theta) \, \hat{\kappa}_<^{lin}(\theta') \, \xi(\theta, \theta') \right]. \tag{35}$$

We recall that $\hat{\gamma} = - \frac{\vartheta_{\hat{\kappa}_<^{lin}}}{2} \frac{d\zeta(\hat{\kappa}_<^{lin}(\theta))}{d\vartheta_{\hat{\kappa}_<^{lin}}}$.

3. SCGF on the reduced shear representation

We want now to extend the SCGF $\varphi^\gamma(\lambda)$ to the SCGF for the observed aperture mass $\hat{M}_{ap}^g$ given in Eq. (24). For this sake, we remember that the (normalized) reduced shear can be expressed as:

$$\hat{g}(\hat{\kappa}_<^{lin}(\theta), w_{\text{eff}}) = \frac{\hat{\gamma}}{1 - \kappa}$$

$$= - \frac{1}{2} \frac{\vartheta \, \hat{\kappa}_<(\vartheta)'}{1 - w_{\text{eff}}[\hat{\kappa}_<(\vartheta) - \frac{1}{2}\vartheta \, \hat{\kappa}_<(\vartheta)']} \tag{36}$$

in terms of the variables ϑ, $\hat{\kappa}_<(\vartheta)$, and the projection factor w_{eff}. The contraction principle allows us to extend

the LDP to any continuous functional of the initial profile, and we can invoke it to replace $\hat{\gamma}$ by $\hat{g}$ in Eq. (35) and obtain the SCGF on the *reduced shear representation*:

$$\varphi^g(\lambda) = - \inf_{\hat{\kappa}_<^{lin}} \left[\lambda \int d\theta \, \tilde{Q}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta)) \, \hat{g}(\hat{\kappa}_<^{lin}(\theta), w_{\text{eff}}) \right.$$

$$\left. + \frac{\sigma_F^2}{2} \int d\theta \, d\theta' \, \hat{\kappa}_<^{lin}(\theta) \, \hat{\kappa}_<^{lin}(\theta') \, \xi(\theta, \theta') \right]. \tag{37}$$

4. Summary

The SCGFs $\varphi^\kappa(\lambda)$ and $\varphi^\gamma(\lambda)$ are two (equivalent) expressions the SCGF for $\hat{M}_{ap}$ given respectively in the convergence and shear representations. The SCGF $\varphi^g(\lambda)$ is the extension of $\varphi^\gamma(\lambda)$ where the reduced shear is used as central observable quantity, and therefore corresponds to the SCGF for the physical $\hat{M}_{ap}^g$.

If the infs in Eqs. (34), (35), and (37) can be computed, the statistical properties of $\hat{M}_{ap}$ and $\hat{M}_{ap}^g$ will follow. It should be stressed that the infimum of a functional doesn't need to agree with a minimum of that functional, but we will only focus here on the determination of minima of our functionals of interest.

IV. PRACTICAL IMPLEMENTATIONS

The actual resolution of the minimization problems obtained in the previous section is not straightforward with no guarantee that it actually converges. We present in the following the different approaches that we effectively employed to solve the minimization problem and check that we have consistent results. Details on the numerical tests can be found in the appendix C.

A. Extremization with the help of the Euler-Lagrange equations

An extremum of the functional $\varphi^\kappa(\lambda)$ given in Eq. (34) can be obtained by imposing

$$\frac{\delta \varphi^\kappa(\lambda)}{\delta \hat{\kappa}_<^{lin}(\theta)} = 0 \,. \tag{38}$$

Indeed, it follows from Eqs. (33), (34), and (38) that the linear profile has to obey the equation:

$$\hat{\kappa}_<^{lin}(\theta) = - \frac{\lambda}{\sigma_F^2} \int d\theta' \sigma^2(\theta, \theta') \frac{\delta}{\delta \hat{\kappa}_<^{lin}} \left[\tilde{V}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta')) \right.$$

$$\left. \times \frac{\partial \zeta(\hat{\kappa}_<^{lin}(\theta'))}{\partial \hat{\kappa}_<^{lin}} \, \hat{\kappa}_<^{lin'}(\theta') \right] \tag{39}$$

where

$$\frac{\delta}{\delta\hat{\kappa}_<^{lin}} = \frac{\partial}{\partial\hat{\kappa}_<^{lin}} - \frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\partial}{\partial\hat{\kappa}_<^{lin\prime}} \tag{40}$$

is the Euler-Lagrange operator. The term inside squared brackets in Eq. (39) has the nice property of being linear on $\hat{\kappa}_<^{lin\prime}$, what leads to great simplifications. Indeed, the use of the explicit form of $\vartheta_{\hat{\kappa}_<^{lin}}$ given in Eq. (28) and the definition of $\tilde{V}$ given in Eq. (32) lead directly to:

$$\hat{\kappa}_<^{lin}(\theta) = \frac{\lambda}{\sigma_F^2}\int \mathrm{d}\theta'\,\sigma^2(\theta,\theta')\Big[\tilde{U}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta'))$$
$$\times \zeta^{-1/2}(\hat{\kappa}_<^{lin}(\theta'))\frac{\partial\zeta(\hat{\kappa}_<^{lin}(\theta'))}{\partial\hat{\kappa}_<^{lin}}\Big]. \tag{41}$$

Also from Eq. (38) it follows that:

$$\frac{\sigma_F^2}{2}\int \mathrm{d}\theta\,\mathrm{d}\theta'\,\hat{\kappa}_<^{lin}(\theta)\,\hat{\kappa}_<^{lin}(\theta')\,\xi(\theta,\theta') = \frac{\lambda}{2}\int \mathrm{d}\theta\mathrm{d}\theta'\hat{\kappa}_<^{lin}(\theta)$$
$$\times \Big[\tilde{U}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta'))\zeta^{-1/2}(\hat{\kappa}_<^{lin}(\theta'))\frac{\partial\zeta(\hat{\kappa}_<^{lin}(\theta'))}{\partial\hat{\kappa}_<^{lin}}\Big] \tag{42}$$

showing that the actual knowledge of $\xi(\theta,\theta')$ assumed in Eq. (30) is not necessary to the solution of the extremization problem.

We obtain, finally,

$$\varphi^\kappa(\lambda) = -\lambda\int \mathrm{d}\theta\,\tilde{V}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta;\lambda))\frac{\partial\zeta(\hat{\kappa}_<^{lin}(\theta;\lambda))}{\partial\hat{\kappa}_<^{lin}}\hat{\kappa}_<^{lin\prime}(\theta;\lambda)$$
$$-\frac{\lambda}{2}\int \mathrm{d}\theta\,\mathrm{d}\theta'\,\hat{\kappa}_<^{lin}(\theta;\lambda)\Big[\tilde{U}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta';\lambda))$$
$$\times \zeta^{-1/2}(\hat{\kappa}_<^{lin}(\theta';\lambda))\frac{\partial\zeta(\hat{\kappa}_<^{lin}(\theta';\lambda))}{\partial\hat{\kappa}_<^{lin}}\Big] \tag{43}$$

where $\hat{\kappa}_<^{lin}(\theta;\lambda)$ is a solution of the integral equation (41) for each value of λ, and $\vartheta_{\hat{\kappa}_<^{lin}}(\theta;\lambda) = \theta\,\zeta^{-1/2}(\hat{\kappa}_<^{lin}(\theta;\lambda))$.

From one side, the linearity of our action in $\hat{\kappa}_<^{lin\prime}$ allows to obtain the simple expression for $\hat{\kappa}_<^{lin}(\theta)$ given in Eq. (41). From the other side, the same linearity of the action in $\hat{\kappa}_<^{lin\prime}$ forbids us of exploring the stability of the extremum through second variations of the action. We will assume that $\hat{\kappa}_<^{lin}(\theta)$ conducts to a maximum of the SCGF on the interval on which it is defined.

The SCGF is given by Eq. (43) as long as Eq. (41) can be solved uniquely for each value of λ. In order to study the existence of unique solutions for this equation, we take the first variation of Eq. (41), obtaining:

$$\frac{\sigma^2(\theta,\theta')}{\sigma_F^2}\frac{\delta}{\delta\hat{\kappa}_<^{lin}(\theta)}\Big[\tilde{U}(\vartheta_{\hat{\kappa}_<^{lin}}(\theta'))$$
$$\times \zeta^{-1/2}(\hat{\kappa}_<^{lin}(\theta'))\frac{\partial\zeta(\hat{\kappa}_<^{lin}(\theta'))}{\partial\hat{\kappa}_<^{lin}}\Big] = \frac{\delta_{\mathrm{D}}(\theta-\theta')}{\lambda} \tag{44}$$

from which we can determine critical values of λ on the solution of Eq. (41).

B. Skewness

Expanding $\varphi(\lambda)$ given in Eq. (31) around $\lambda = 0$ leads to the expression of cumulants. For the skewness we have [18]:

$$\hat{s}_3^\kappa = 3\nu_2\frac{\int \mathrm{d}x\,\tilde{U}(x)\,\Sigma^2(x)}{[\int \mathrm{d}x\,\tilde{U}(x)\,\Sigma(x)]^2} + 3\frac{\int \mathrm{d}x\,x\,\tilde{U}(x)\,\Sigma(x)\,\Sigma'(x)}{[\int \mathrm{d}x\,\tilde{U}(x)\,\Sigma(x)]^2} \tag{45}$$

with

$$\Sigma(x) = \int \mathrm{d}y\,\sigma^2(x,y)\,\tilde{U}(y). \tag{46}$$

(Note that the coefficient in front of the second term in Eq. (45) is $6/d$, where d is the dimension of space in which spherical collapse is considered). It is worth to recall that the reduced skewness of the aperture mass is then

$$s_3^\kappa = \frac{1}{w_{\mathrm{eff}}}\hat{s}_3^\kappa \tag{47}$$

We can do the same exercise by expanding $\varphi^g(\lambda)$ around $\lambda = 0$. At lowest order, the correction on the skewness introduced by the use of the reduced shear is:

$$\hat{s}_3^g - \hat{s}_3^\kappa = 6\,w_{\mathrm{eff}}\frac{\int \mathrm{d}x\,\tilde{Q}(x)\frac{x}{2}\Sigma'(x)\left(\Sigma(x)+\frac{x}{2}\Sigma'(x)\right)}{[\int \mathrm{d}x\,\tilde{U}(x)\,\Sigma(x)]^2} \tag{48}$$

and

$$s_3^g - s_3^\kappa = \frac{1}{w_{\mathrm{eff}}}\left(\hat{s}_3^g - \hat{s}_3^\kappa\right). \tag{49}$$

As expected the expression of s_3^κ and s_3^g depend on the choice of power spectrum and for a sake of simplicity we illustrate our results for power law spectra. More specifically we can analyse the relative importance of the correction term given in Eq. (48) by computing the ratio $(s_3^g - s_3^\kappa)/(w_{\mathrm{eff}}s_3)$ which is then independent of w_{eff}. As shown in Fig. 3, the impact on the skewness of the distribution of $\hat{M}_{ap}^g$ has a few percent deviation from the skewness for $\hat{M}_{ap}$ as function of the spectral index n.

C. The SCGF and the rate functions

The Euler-Lagrange equations presented in Sec. IV A can be used to obtain φ^κ but the action becomes considerably more involved when we are interested in φ^g. We treated the problem numerically in two alternative ways: the implementation of Eqs. (41) and (43), or the direct extremization methods available in Mathematica [32] such as the FindMinimum routine. As shown in appendix C, direct extremization and the Euler-Lagrange methods agree to high precision on the range of λ of our interest and for the discretizations of the interval considered.
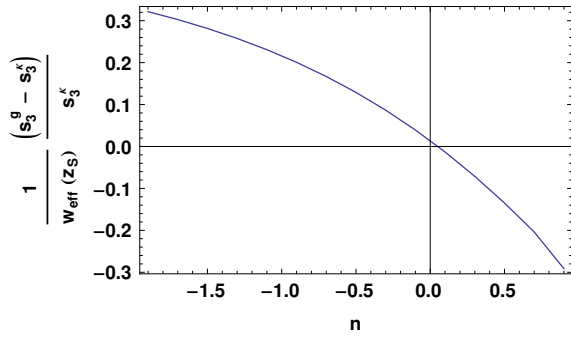
FIG. 3: Deviation of the skewness for $\hat{M}_{ap}^g$ and $\hat{M}_{ap}$ normalized by w_{eff}. Since $0 \leq w_{\text{eff}} \leq 0.3$ in Einstein-de Sitter for $z_S = 1$, we see that the highest possible correction to the skewness in this case is of order of 10%.

The general output from the numerical solutions can be seen in Fig. 4. On this plot $\varphi^\kappa(\lambda)$ is computed using the Euler-Lagrange strategy and $\varphi^\gamma(\lambda)$ and $\varphi^g(\lambda)$ derive from direct extremization. We should remark that although displayed in two different representations we have $\varphi^\kappa(\lambda) = \varphi^\gamma(\lambda)$, and this is recovered in the numerical reconstructions (see Sec. C 1 for more details). We note that moving from the shear to the reduced shear induces some changes on the SCGF and also moves the critical points closer to the origin. The critical points for φ^κ can be also obtained from Eq. (44).
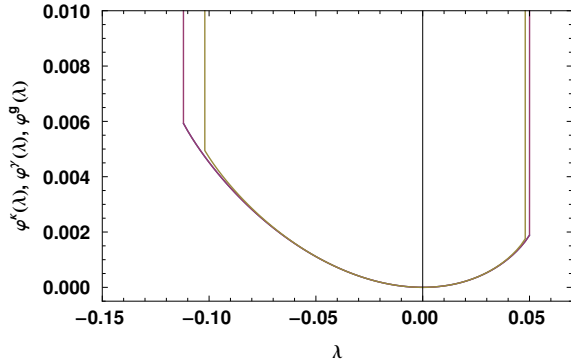

FIG. 4: The SCGFs $\varphi^\kappa(\lambda)$, $\varphi^\gamma(\lambda)$, and $\varphi^g(\lambda)$ for $n = -1.5$. The curves for $\varphi^\kappa(\lambda)$ and $\varphi^\gamma(\lambda)$ coincide as they should. The SCGF for $\hat{M}_{ap}^g$ is however slightly different and has different critical points. We use $w_{\text{eff}} = 0.1$ here.

The SCGF is always a convex function, the same is not true for the rate function. As already noted, and discussed in appendix B, SCGF and rate function can be obtained from each other by Legendre transform as long as the rate function is convex. If the rate function ceases to be convex, the SCGF will still be given by the Legendre-Fenchel transform of the rate function, but will present points of non-differentiability. We then note that the Legendre-Fenchel transform of such a SCGF will produce the convex envelope of the rate function only, and

not the rate function itself[7].

D. The One-point PDF

We complete our investigations by evaluating the impact of those changes on the shape of the one-point PDF of the $\hat{M}_{ap}$ values. The calculation is based on the computation of the inverse Laplace transform of the cumulant generating function (see for instance [17] for details). For a sake of simplicity we perform this calculation using an effective form for the SCGF (as in [20]). Such an effective form is based on an effective vertex generating function $\zeta_{\text{eff}}(\hat{\kappa}_<^{lin})$ satisfying:

$$\varphi_{\text{eff}}(\lambda) = \lambda\, \zeta_{\text{eff}}(\hat{\kappa}_<^{lin}) - \frac{1}{2}\lambda\, \hat{\kappa}_<^{lin}\, \zeta_{\text{eff}}'(\hat{\kappa}_<^{lin}) \quad (50a)$$

$$\hat{\kappa}_<^{lin} = \lambda\, \zeta_{\text{eff}}'(\hat{\kappa}_<^{lin})\,. \quad (50b)$$

where $\zeta_{\text{eff}}(\hat{\kappa}_<^{lin})$ is adjusted so that $\varphi_{\text{eff}}(\lambda)$ provides a good fit to the SCGF we computed, in particular reproducing its critical behaviors. In practice one can get a very good fit with a fifth order polynomial for ζ_{eff}.
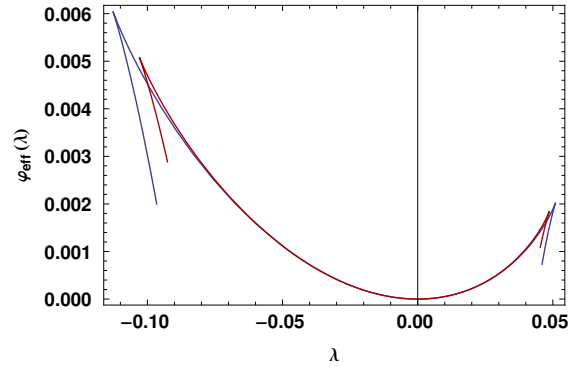

FIG. 5: The effective cumulant generating functions $\varphi_{\text{eff}}^\kappa$ and φ_{eff}^g satisfying Eq. (50). The projection factor $w_{\text{eff}} = 0.1$ is used on the φ_{eff}^g data.

The effective cumulant generating function obtained this way reproduces extremely accurately the global be-

---

[7] A duality property connects features of the SCGF and the convex envelope of the rate function [15]: if the SCGF goes to infinity at critical points, then the convex envelope of the rate function will have segments of constant derivative beyond critical points. The location of the critical points of the SCGF give the inclination of the affine segments of the rate function, while the lateral derivatives of the SCGF at the critical points give the location of its critical points. We can understand from the LDP that linear segments on the rate function implies exponential decay for the probability density function. Unless we know the geometry of the space of solutions for the minimization problem, we cannot know the rate function globally for our problem. All we can access from our method is therefore the segment that corresponds to the non-affine segment of the convex envelope just described.

havior obtained previously in particular for its critical points.

The reconstruction of the one point PDF of $\hat{M}_{ap}$ is then obtained from the following form,

$$P(\hat{M}_{ap}) = \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\lambda}{2\pi} \exp[-\lambda\,\hat{\kappa}_< + \varphi_{\hat{\kappa}_<}(\lambda)]. \qquad (51)$$

where the function $\varphi_{\hat{\kappa}_<}(\lambda)$ is built from the SCGF,

$$\varphi_{\hat{\kappa}_<}(\lambda) = \frac{1}{\hat{\sigma}^2}\varphi^\kappa(\lambda\,\hat{\sigma}^2) \qquad (52)$$

in such a way that $\hat{\sigma}^2$ matches the expected variance of $\hat{M}_{ap}$. The actual computation of such inverse Laplace transforms has been described in referenced papers and is based on the integration along the imaginary axis.
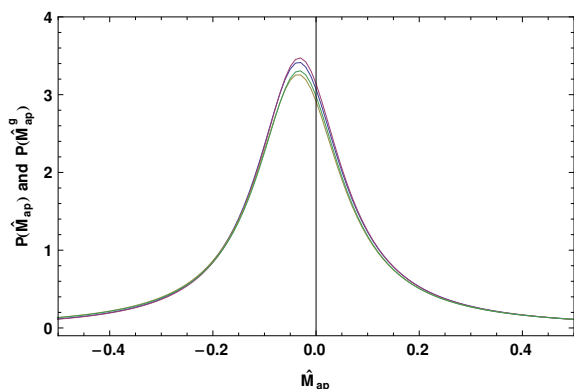

FIG. 6: PDFs obtained from the inverse Laplace transformation of φ^κ and φ^g for $\hat{\sigma} = 0.4$ (top two curves) and $\hat{\sigma} = 0.7$ (bottom two curves). The $\hat{M}_{ap}^g$ is reconstructed for $w_{\mathrm{eff}} = 0.1$. In each case $P(\hat{M}_{ap}^g)$ is sightly larger than $P(\hat{M}_{ap})$ for $\hat{M}_{ap} \approx 0$ exhibiting slightly stronger non-Gaussianities.
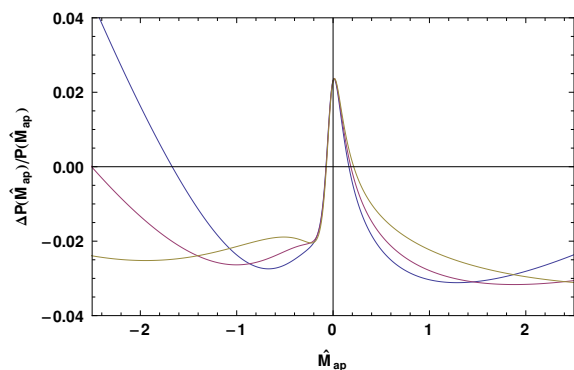

FIG. 7: Difference of the PDFs shown on Fig. 6 for $\hat{\sigma} = 0.4$ (blue curve), $\hat{\sigma} = 0.5$ (red curve) and $\hat{\sigma} = 0.7$ (yellow curve).

The resulting PDFs, $P(\hat{M}_{ap})$ and $P(\hat{M}_{ap}^g)$ built respectively from the shear field and the reduced shear field are shown in Fig. 6. For these ranges of $\hat{M}_{ap}$ and this value of w_{eff} the relative errors are about a few percent as

shown on Fig. 7 consistent with our findings concerning the skewness. What these results show however is that the extra non-linearities contained in the reduced shear expression have little effects on the global shape of the PDFs.

V. CONCLUSION

In this paper we present the derivation of the cumulant generating function and the corresponding one-point PDF of the aperture mass, $\hat{M}_{ap}$, in a general framework. In particular we take into account that aperture mass can only be built in practice from the measured reduced shear (=shear/(1-convergence)) and will be built in general from a variety of filter shapes. So, although it is still not a realistic prediction since we did not take into account the exact conical geometry of the observations, our result extends previous computations in two fundamental ways: regarding the shape of the filter and by lifting the usual identification of the shear and the reduced shear.

On a more fundamental level it shows that it is now possible to explore a whole new class of problems, namely the derivation of statistical properties of quantities obtained by general (symmetric) filtering of functionals of the density profiles[8]. That alone illustrates the reach of the Large Deviation Principle formalism and the fact that it can be exploited in much more elaborate situations than previously thought from direct PT resummations (as pioneered in [26, 33] and extended in [17]).

What we found is that the results that were previously obtained, specifically in [20], are robust with respect to these extensions. In that study the compensated filter was indeed assumed to be simply built from the difference of two top-hat filters applied to the convergence field. Our construction is much more elaborate. We find however that the general properties of the cumulant generating functions are left unchanged: they exhibit for instance critical values for both positive and negative values of the parameters whether the $\hat{M}_{ap}$ is built from the shear of the reduced shear. Differences can be noted however regarding the position of the critical points: as expected the critical values are closer to the origin as large excursions in κ lead to stronger non-linear effects. The impact of this effect is however rather mild. We have expressed it in terms of the $\hat{M}_{ap}$ skewness and the $\hat{M}_{ap}$ one-point PDF.

The differences we found are of the few percent range and would require therefore extremely good measurements to be of significant impact.

---

[8] We note that finding a profile satisfying a given condition can be re-expressed in terms of a condition on a non-Markovian random walk. Whether it has a connection to the computation of halo mass functions in a Press and Schechter like approach is largely an open problem.

Formally the construction of the scaled cumulant generating function is based on the minimization of the a functional form of the density in concentric cells, that is of the whole density profile once we are in the continuous limit. In our study we did not explore in much details how to do such minimization efficiently but checked that our results were correct from different approaches (that are equivalent in the continuous limit) leaving room for further improvements.

[1] D. M. Wittman, J. A. Tyson, D. Kirkman, I. Dell'Antonio, and G. Bernstein, Nature (London) **405**, 143 (2000), astro-ph/0003014.

[2] L. Van Waerbeke *et al.*, A&A**358**, 30 (2000), astro-ph/0002500.

[3] D. J. Bacon, A. R. Refregier, and R. S. Ellis, MN-RAS**318**, 625 (2000), astro-ph/0003008.

[4] L. Van Waerbeke and Y. Mellier, ArXiv Astrophysics e-prints (2003), astro-ph/0305089.

[5] D. Munshi, P. Valageas, L. Van Waerbeke, and A. Heavens, Physics Reports **462**, 67 (2008).

[6] M. Kilbinger, Reports on Progress in Physics **78**, 086901 (2015).

[7] L. Fu *et al.*, A&A**479**, 9 (2008), 0712.0884.

[8] M. R. Becker *et al.*, Phys. Rev. D **94**, 022002 (2016), 1507.05598.

[9] L. Fu *et al.*, MNRAS**441**, 2725 (2014), 1404.5469.

[10] T. Kacprzak *et al.*, MNRAS**463**, 3653 (2016), 1603.05040.

[11] S. R. S. Varadhan, *Large Deviations and Applications* (SIAM, 1984).

[12] F. den Hollander, *Large Deviations* (AMS, 2008).

[13] R. S. Ellis, The theory of large deviations and applications to statistical mechanics, Lecture Notes for École de Physique Les Houches, 2009.

[14] S. R. S. Varadhan, Large deviations, https://math.nyu.edu/faculty/varadhan/LDP/1-2.pdf, 2010.

[15] H. Touchette, Physics Reports**478**, 1 (2009), 0804.0327.

[16] P. Valageas, A&A**382**, 412 (2002).

[17] F. Bernardeau, C. Pichon, and S. Codis, Physical Review D **90**, 103519 (2014).

[18] F. Bernardeau and P. Reimberg, Phys. Rev. **D94**, 063520 (2016), 1511.08641.

[19] C. Uhlemann, S. Codis, C. Pichon, F. Bernardeau, and P. Reimberg, Monthly Notices of the Royal Astronomical Society **460**, 1529 (2016).

[20] F. Bernardeau and P. Valageas, Astron. Astrophys **364**, 1 (2000).

[21] D. Munshi, P. Valageas, and A. J. Barber, Monthly Notices of the Royal Astronomical Society **350**, 77 (2004).

[22] P. Valageas, A. J. Barber, and D. Munshi, Monthly Notices of the Royal Astronomical Society **347**, 654 (2004).

[23] A. J. Barber, D. Munshi, and P. Valageas, Monthly Notices of the Royal Astronomical Society **347**, 667 (2004).

[24] N. Kaiser, The Astrophysical Journal **439**, L1 (1995).

[25] P. Schneider, Monthly Notices of the Royal Astronomical Society **283**, 837 (1996).

[26] F. Bernardeau, A&A**291**, 697 (1994).

[27] F. Bernardeau, Astrophys. J. **433**, 1 (1994).

[28] F. Bernardeau, Astron. Astrophys. **301**, 309 (1995), astro-ph/9502089.

[29] R. Juszkiewicz, F. R. Bouchet, and S. Colombi, ApJL**412**, L9 (1993), arXiv:astro-ph/9306003.

[30] F. Bernardeau, L. Van Waerbeke, and Y. Mellier, Astron. Astrophys. **322**, 1 (1997), astro-ph/9609122.

[31] R. G. Crittenden, P. Natarajan, U.-L. Pen, and T. Theuns, The Astrophysical Journal **568**, 20 (2002).

[32] Wolfram Research, Inc., Mathematica 9.0.

[33] F. Bernardeau, Astrophys. J. **392**, 1 (1992).

[34] S. W. Hawking and G. F. R. Ellis*The large scale structure of space-time* Vol. 1 (Cambridge university press, 1973).

[35] P. H. Reimberg and L. R. Abramo, Classical and Quantum Gravity **30**, 065020 (2013).

[36] M. Bartelmann, Classical and Quantum Gravity **27**, 233001 (2010).

[37] P. Peter and J.-P. Uzan, *Primordial cosmology* (Oxford University Press, 2013).

[38] V. Perlick, Classical and Quantum Gravity **13**, 529 (1996).

[39] T. Schramm and R. Kayser, A&A**299**, 1 (1995), astro-ph/9408064.

[40] N. Straumann, *General relativity and relativistic astrophysics* (Springer Science & Business Media, 2012).

[41] S. R. S. Varadhan, *Probability Theory* (AMS, 2001).

Appendix A: Lensing in Cosmology

Null vectors in Lorentzian geometry are orthogonal to themselves: let γ be a null geodesic, then $\langle \gamma', \gamma' \rangle = 0$ (here $\langle \cdot, \cdot \rangle$ means contraction with the spacetime metric, and the prime indicated derivative with respect to the affine parameter). Given γ', a four dimension orthogonal basis at each point will be completed by another null vector orthogonal to γ' and two space-like vectors. The geometry of a thin light beam will be described by γ', that gives its direction of propagation, and by a vector orthogonal to it describing the distance of each null geodesic on the beam. Removing the second null vector by taking the quotient by an equivalence relation [34], the separation of the geodesics with respect to the fiducial one is described in terms of the so called *Jacobi fields*, i.e., vectors on a two-dimensional space-like screen. The history of the light beam's deformations can be told by the evolution of its sections by successive screens, and the

differential equation governing this geometrical evolution is known as *Jacobi equation*:

$$\mathcal{D}'' + \mathcal{T}\mathcal{D} = 0 \,. \tag{A1}$$

The object $\mathcal{T}$ is the *optical tidal matrix* that depends on the Ricci and Weyl tensors:

$$\mathcal{T} = \begin{pmatrix} \mathcal{R} + \mathrm{Re}\mathcal{F} & \mathrm{Im}\mathcal{F} \\ \mathrm{Im}\mathcal{F} & \mathcal{R} - \mathrm{Re}\mathcal{F} \end{pmatrix} \,, \tag{A2}$$

where, in components, $\mathcal{R} = \frac{1}{2}R_{\mu\nu}k^\mu k^\nu$ and $\mathcal{F} = \frac{1}{2}\epsilon^\alpha \epsilon^\beta C_{\alpha\mu\beta\nu}k^\mu k^\nu$ (if E_1, E_2 are the orthonormal basis vectors on the space-like screen, then $\epsilon = E_1 + \mathrm{i}E_2$. Also $k = \gamma'$). The initial conditions to be provided to Eq. (A1) are $\mathcal{D}(0) = 0$, $\mathcal{D}'(0) = \mathbf{1}$. The transport problem is solved once one knows $\mathcal{D}$ for all values of the affine parameter [35].

We want here solve this problem on a perturbed Friedmann-Lemaitre-Robertson-Walker (FLRW) space-time. Let ϕ be the gravitational potential that parametrizes the scalar perturbations on FLRW. If $\phi = 0$ the Weyl tensor vanishes and the optical tidal matrix is $\mathcal{T}_0 = K\mathbf{1}$, where K is the sectional curvature of the space-time. The solution of Eq. (A1) in this case is $\mathcal{D}(\chi) = \mathcal{D}_0(\chi)\mathbf{1}$, with

$$\mathcal{D}_0(\chi) = \frac{c/H_0}{\sqrt{1 - \Omega_\mathrm{m} - \Omega_\Lambda}} \sinh\left(\sqrt{1 - \Omega_\mathrm{m} - \Omega_\Lambda} \frac{H_0\chi}{c} \right) \tag{A3}$$

that corresponds to the Jacobi fields in spaces of constant curvature, or the comoving angular diameter as function of the comoving radial distance χ.

To take into account the contributions of the gravitational potential ϕ we should go at least to the first order correction on $\mathcal{T}_1$, which includes the Ricci and Weyl tensors associated to the linear perturbations on the metric. We will assume that the lens effect is computed along a null geodesic of the unperturbed space-time, what corresponds to the *Born approximation*. It can be shown that the first order correction on $\mathcal{D}$ is [36, 37]:

$$\left(\frac{\mathcal{D}_1}{\mathcal{D}_0} \right)_{ab} (\boldsymbol{\theta}, \chi) = -\nabla_a \nabla_b \phi_L(\boldsymbol{\theta}, \chi) \tag{A4}$$

where $a, b = 1, 2$, and ϕ_L is the *lens potential*:

$$\phi_L(\boldsymbol{\theta}, \chi) = \frac{2}{c^2} \int_0^\chi \mathrm{d}\chi' \frac{\mathcal{D}_0(\chi - \chi')\mathcal{D}_0(\chi')}{\mathcal{D}_0(\chi)} \phi(\mathcal{D}_0(\chi')\boldsymbol{\theta}, \chi') \,. \tag{A5}$$

The deformation of the beam's sections relative to its spreading on unperturbed spacetime is:

$$\frac{\mathcal{D}}{\mathcal{D}_0} = \mathbf{1} + \frac{\mathcal{D}_1}{\mathcal{D}_0} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \tag{A6}$$

where we make a decomposition in trace and traceless contributions by defining *convergence* $\kappa := \frac{1}{2}(\nabla_1 \nabla_1 +$

$\nabla_2 \nabla_2)\phi_L$ and *shear* $\boldsymbol{\gamma} := (\gamma_1, \gamma_2)$, with $\gamma_1 = \frac{1}{2}(\nabla_1 \nabla_1 - \nabla_2 \nabla_2)\phi_L$, and $\gamma_2 = \nabla_1 \nabla_2 \phi_L$.

The determinant of the matrix presented in Eq. (A6) vanishes at points corresponding exactly to conjugate points along the fiducial geodesic. Since the presence of conjugate points along the fiducial geodesic is a sufficient condition for multiple imaging [38], the vanishing of the determinant roughly indicates the transition from weak to strong lensing.

1. Lens Map and reduced shear

The *lens map* is the map whose differential is the deformation matrix given in Eq. (A6). Instead of writing vectors in $\mathbb{R}^2$, we will use here complex objects. Let $z = x + \mathrm{i}y \in \mathbb{C}$ and $f(z) = u(x,y) + \mathrm{i}v(x,y)$ a function of z. Using this notation we conclude that:

$$f(z) = u(x,y) + \mathrm{i}v(x,y) := \left(x - \frac{\partial \phi_L}{\partial x} \right) + \mathrm{i}\left(y - \frac{\partial \phi_L}{\partial y} \right) \tag{A7}$$

is the lens map. It follows from the definitions of shear and convergence that:

$$\frac{1}{2}\left[\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + \mathrm{i}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) \right] = 1 - \kappa \tag{A8a}$$

$$\frac{1}{2}\left[\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right) + \mathrm{i}\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) \right] = -\gamma \tag{A8b}$$

where we write $\gamma = \gamma_1 + \mathrm{i}\gamma_2$. This system of equations is equivalent to Beltrami's equation with Beltrami coefficient $\gamma/(1 - \kappa)$ [39], and the lens map is therefore quasi-conformal. We observe that if $\gamma = 0$, then Eqs. (A8) are the Cauchy-Riemann equations.

Eqs. (A8) imply that infinitesimal ellipses are transformed into circles by the lens map. Indeed,

$$p(x,y) := \frac{1 + \frac{|\gamma|}{1 - \kappa}}{1 - \frac{|\gamma|}{1 - \kappa}} \tag{A9}$$

is the ratio of the axes of that ellipse. The quantity

$$g := \frac{|\gamma|}{1 - \kappa} \tag{A10}$$

defines the *reduced shear*.

The argument of the major axis of the infinitesimal ellipses allows to determine the phase of the Beltrami coefficient [40].

Appendix B: Large Deviation Theory

Convergence (or divergence) is one of the most central, and most studied concepts in mathematics. In probability theory different kinds of convergence can be defined.

A sequence of random variables X_n converges *strongly* to X if $\Pr[\lim_{n\to\infty} X_n = X] = 1$. Alternatively, a sequence is said to *converge in probability* to a given element X if $\lim_{n\to\infty} \Pr[|X_n - X| \geq \varepsilon] = 0$, for any $\varepsilon > 0$ given. We can define also the *weak convergence* (or convergence in law) by saying that a sequence of probability measures α_n converges weakly to a limiting probability measure α if

$$\lim_{n\to\infty} \int f(x)\mathrm{d}\alpha_n = \int f(x)\mathrm{d}\alpha$$

for every bounded function $f(x)$ on $\mathbb{R}$. Equivalently, if ϕ_n and ϕ are respectively the characteristic functions of α_n and α, $\lim_{n\to\infty} \phi_n(t) = \phi(t)$. The portmanteau theorem in probability theory [41] states that if α_n converges weakly to α on $\mathbb{R}$, and $C \in \mathbb{R}$ is a closed set, then

$$\limsup_{n\to\infty} \alpha_n(C) \leq \alpha(C)$$

while for open sets $G \in \mathbb{R}$,

$$\liminf_{n\to\infty} \alpha_n(G) \geq \alpha(G)\,.$$

If $A \in \mathbb{R}$ is such that $\alpha(\bar{A} - A^o) = 0$ ($\bar{A}$ is the closure of A and A^o its interior), then

$$\lim_{n\to\infty} \alpha_n(A) = \alpha(A)\,.$$

We may think of a class of problems for which a typical value exists, and events far from this typical value will be classified as *rare*. This would be the case for a process described by a probability density function with exponentially decaying tails. Events on the tails are the prototype of rare events. The definition of *large deviation principle* builds on the portmanteau theorem to include the idea that probability measure associated to rare events are exponentially suppressed, and introduces the *rate function* that modulates the exponential decay.

We say that $\{P_\varepsilon\}$ obeys the Large Deviation Principle (LDP) with a rate function I if there exists a function $I(\cdot) : \mathbb{R} \to [0, \infty]$ lower semicontinuous with compact level sets such that:

i) For each closed set $C \in \mathbb{R}$

$$\limsup_{\varepsilon\to 0} \varepsilon \log P_\varepsilon(C) \leq -\inf_{x\in C} I(x)$$

ii) For each open set $G \in \mathbb{R}$

$$\liminf_{\varepsilon\to 0} \varepsilon \log P_\varepsilon(G) \geq -\inf_{x\in G} I(x)$$

If $\inf_{x\in A^o} I(x) = \inf_{x\in A} I(x) = \inf_{x\in \bar{A}} I(x)$, then

$$\lim_{\varepsilon\to 0} \varepsilon \log P_\varepsilon(A) = -\inf_{x\in A} I(x)\,. \tag{B1}$$

The parameter ε has to be identified as some limiting parameter on each problem of interest. For collections of identical, identically distributed random variables (i.i.d.), for instance, $\varepsilon = 1/n$. This definition is general enough to allow Large Deviations Theory to be applied to a large variety of problems of different levels of abstraction, and in general we can replace $\mathbb{R}$ for any complete separable metric space ($\mathbb{R}^n$, for instance) on the definition. We can rephrase the LDP in terms of a family of random variables $\{X_i\}$, by writing $\lim_{\varepsilon\to 0} \varepsilon \log P_\varepsilon(\{X_i\} \in A) = -\inf_{x\in A} I(x)$, where we concentrate on Eq. (B1) because it can be taken as the definition of the Large Deviation Principle for our needs.

We may ask now two questions: $i)$ are there families of random variables that obey the LDP? $ii)$ What are the most immediate consequences of LDP? To answer the first question, we can quote the famous Cramer's theorem [13]: Let $\{X_i\}$ be a sequence of i.i.d. random vectors on $\mathbb{R}^n$, and $S_k/k := \sum_{i=1}^{k} X_i/k$ its the *sample mean*. The sequence of sample means S_k/k satisfies the LDP with rate function $I(x) = \sup_{\lambda\in\mathbb{R}^n}[\lambda x - c(\lambda)]$, where $c(\lambda) = \log \mathbb{E}[e^{\lambda x}]$ is the cumulant generating function. To have an idea of the origin of the theorem, take $x > \mathbb{E}[X_1]$, $\lambda > 0$,

$$
\begin{aligned}
\Pr(S_k/k \in [x, \infty)) &= \Pr(\lambda S_k \geq k\lambda x) \\
&\leq \exp[-k\lambda x]\mathbb{E}[e^{\lambda S_k}] \\
&= \exp[-k\lambda x]\prod_{i=1}^{k} \mathbb{E}[e^{\lambda X_i}] \\
&= \exp[-k\lambda x]\left(\mathbb{E}[e^{\lambda X_1}]\right)^k \\
&= \exp[-k\lambda x + k\log\mathbb{E}[e^{\lambda X_1}]] \\
&= \exp[-k(\lambda x - c(\lambda))] \tag{B2}
\end{aligned}
$$

Where we use Chebyshev's inequality, independence of X_i and the fact that they are identically distributed to derive this upper bound. This argument does not demonstrate the theorem, but gives a simple illustration of its origins. It points out that rate function and cumulant generating function are related by Legendre-Fenchel transform.

The requirement of independence can be weakened for gaussian random variables. Indeed, for a vector x with mean μ and covariance matrix Σ, the rate function is

$$I(x) = \frac{1}{2}\sum_{ij}(x - \mu)_i\, \Xi_{ij}\, (x - \mu)_j \tag{B3}$$

where we denote $\Xi = \Sigma^{-1}$.

Going back to the second question, we will list two consequences of the LDP: the *contraction principle* and the *Varadhan's lemma*. The contraction principle affirms that the image under a continuous map F of families of random variables satisfying the LDP will also satisfy the LDP with rate function:

$$J(y) = \inf_{x:F(x)=y} I(x)\,. \tag{B4}$$

If F is not injective, the contraction principle encodes the idea that "any large deviation is done in the least

unlikely of all unlikely ways" [12]. If F is a bijection, on the other hand, $J(y) = I(F^{-1}(y))$. The direct verification of the definition of LDP for a given sequence of probability measures (or random variables) may be prohibitively complicated, and therefore the contraction principle is a powerful tool.

The second important consequence of the LDP that we want to list is the Varadhan's lemma: Consider a family of random variables $\{X_i^\varepsilon\}$, $i = 1, \ldots, n$ as a vector in $\mathbb{R}^n$ whose components are X_i^ε. If this family of random variables satisfy the LDP with rate function $I(\cdot)$, and $F : \mathbb{R}^n \to \mathbb{R}$ is bounded, then

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{E}\left[\exp\left(\frac{F(X^\varepsilon)}{\varepsilon}\right)\right] = \sup_{x \in \mathbb{R}^n}[F(x) - I(x)] \quad \text{(B5)}$$

To have an idea of why is this so, we should remember that, because of the LDP, $\mathbb{E}\left[\exp\left(\frac{F(X^\varepsilon)}{\varepsilon}\right)\right] \approx \exp\left[\sup_x [F(x) - I(x)]/\varepsilon\right]$, and use the Laplace method to give an approximative answer to the integral. At the limit $\varepsilon \to 0$ we have the equality. If we consider the simple case where $F(X^\varepsilon) = \sum_i \lambda_i X_i^\varepsilon$, (i.e., the scalar product of X^ε with a vector $\lambda \in \mathbb{R}^n$) we recognize

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{E}\left[\exp\left(\frac{\sum_i \lambda_i X_i^\varepsilon}{\varepsilon}\right)\right] =: \varphi(\lambda) = \sup_{x \in \mathbb{R}^n}[\lambda x - I(x)] \quad \text{(B6)}$$

that is the *scaled cumulant generating function* (SCGF). All the (scaled) cumulants can be obtained from $\varphi(\lambda)$ by partial differentiation.

The Eq. (B6) also states that the SCGF and the rate function are convex conjugates of each other, or that are related by a *Legendre-Fenchel* transformation. The Legendre-Fenchel transformation reduces to the classical Legendre transformation when the supremum is realized on a maximum of the function under consideration. On the set of convex functions the convex transformation is an involution, and therefore not only the SCGF can be obtained from the rate function by convex transformation, but also the rate function can be obtained from the SCGF. If the rate function is not globally convex, the SCGF is still the Legendre-Fenchel transform of the rate function, but only the convex envelope of the rate function can obtained from the SCGF. Indeed, points of non-differentiability of the SCGF will be associated to points where the rate function loses convexity.

Appendix C: Numerical evaluations

Our search for the scaled cumulant generating functions for the aperture mass and physical aperture mass conducted us to φ^κ, φ^γ, and φ^g given in Eqs. (34), (35), (37) respectively. We must now obtain solutions for the minimization problems.

One possible strategy to approach the solution of the minimization problem is presented in Sec. IV A. The relative simplicity of the functional conducts to $\varphi^\kappa(\lambda)$ in Eq. (43), once on solves Eq. (41) for each value of λ in a range of interest. We will refer to it as *Euler-Lagrange (EL) strategy*.

A second possible path is the direct search for minima using the numerical algorithms such as FindMinimum in Mathematica [32], that we will call *Direct Extremization (DE) strategy*.

On both cases we will look for numerical solutions initially close to the linear extrapolation $(\hat{\kappa}_<^{lin})_{init}(\theta) = \frac{1}{\sigma_F^2} \int d\theta' \, \sigma^2(\theta, \theta') \tilde{U}(\hat{\kappa}_<^{lin}(\theta'))$.

We should also remark that the non-linear density given in Eq. (11) is singular at $\tau_{2D} = \nu$. The numerical search for solutions may not find good converging paths because of this divergence and in order to circumvent this limitation, we propose a regularization to the non-linear density:

$$\zeta^{reg}(\tau_{2D}) = \frac{(1+\epsilon)^{\nu/2}}{\left(1 - \left(\frac{\tau_{2D}}{\nu}\right)^2 + \epsilon\right)^{\nu/2}} \quad \text{(C1)}$$

where ϵ is a small parameter (we choose $\epsilon = 0.00001$).

1. Comparison between Euler-Lagrange and Direct Extremization strategies for φ^κ

We start comparing the performances of the Euler-Lagrange and Direct Extremization strategies for the reconstruction of the same object $\varphi^\kappa(\lambda)$.

a. The SCGF

In order to compare the two numerical strategies on the solution of the equation (34), we take the difference of the SCGFs obtained by the EL and DE strategies, as displayed in Fig. 8. The small departure of the solutions validate the DE strategy.
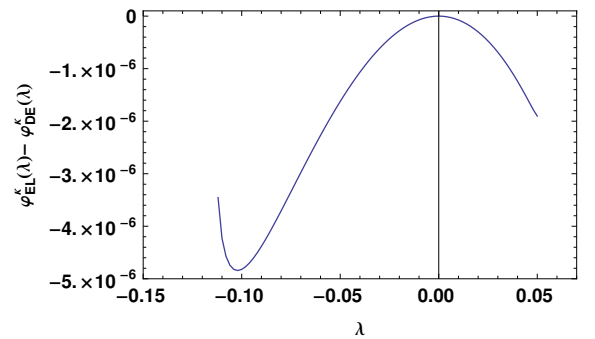

FIG. 8: $\varphi_{EL}^\kappa(\lambda) - \varphi_{DE}^\kappa(\lambda)$ for $n = -1.5$.

b. Skewness

Given the numerical solution $\varphi^\kappa_{EL}(\lambda)$, we can compute numerically the skewness and compare with the theoretical prediction given in Eq. (45).
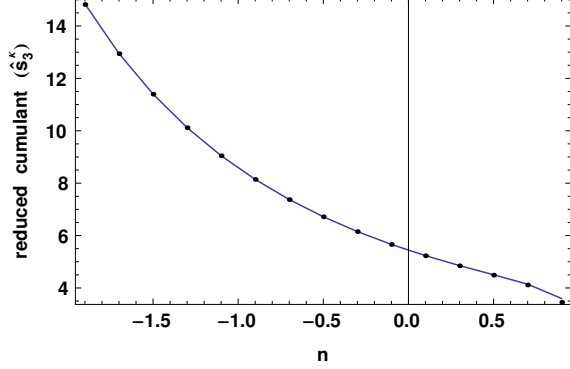

FIG. 9: Theoretical prediction from Eq. (45) and numerical values for the reduced skewness as function of the spectral index.

We can also compare the performance of the EL and DE strategies on the calculation of the skewness on Fig. 10.
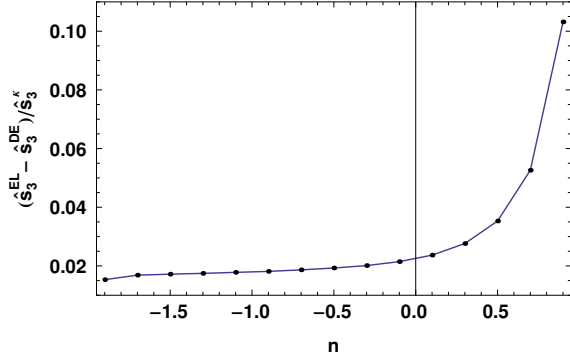

FIG. 10: The difference of the skweness obtained in EL and DE strategies divided by the theoretical prediction given in Eq. (45). We see that the DE tends to give systematically smaller values for the cumulants for higher values of the spectral index.

The DE strategy has the tendency to copy the initial linear guess on small neighborhoods of the origin, impacting the calculation of the derivatives of the SCGF at the origin. This explains the 1% to 10% error.

2. The skewness for $\hat{M}^g_{ap}$

If the use of two different strategies was possible for solving $\varphi^\kappa(\lambda)$, the same is no longer true if we want to obtain $\varphi^g(\lambda)$. In this case the functional is no longer linear on $\hat{\kappa}^{lin'}_<$ and the simplifications obtained on the Sec. IV A have no parallel. We are constrained in this case to use direct extremization.

We can verify the numerical values to the theoretical prediction given in Eq. (48). For this sake we compute $\hat{s}^g_3 - \hat{s}^\kappa_3$ using the DE strategy for $w_{\text{eff}} = 0.1$. The result is shown in Fig. 11. As discussed in Fig. 10, the DE strategy induces to $1-10\%$ error on the skweness, mainly for higher values of n, what also impacts the comparison between numerical points and theoretical prediction in Fig. 11.
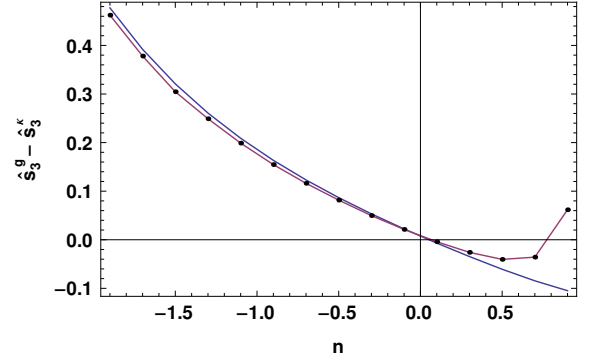

FIG. 11: Theoretical prediction from Eq. (48) and numerical values for the difference of reduced skewness $\hat{s}^g_3 - \hat{s}^\kappa_3$ as function of the spectral index for $w_{\text{eff}} = 0.1$.