



HAL
open science

Large deviation principle at work: Computation of the statistical properties of the exact one-point aperture mass

Paulo Reimberg, Francis Bernardeau

► **To cite this version:**

Paulo Reimberg, Francis Bernardeau. Large deviation principle at work: Computation of the statistical properties of the exact one-point aperture mass. *Physical Review D*, 2018, 97 (2), pp.023524. 10.1103/PhysRevD.97.023524 . hal-01703564

HAL Id: hal-01703564

<https://hal.science/hal-01703564>

Submitted on 8 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Large deviation principle at work: Computation of the statistical properties of the exact 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



(Received 12 September 2017; published 25 January 2018)

We present a formalism based on the large deviation principle (LDP) applied to cosmological density fields, and more specifically to the arbitrary functional of density profiles, and we apply it to the derivation of the cumulant generating function and one-point probability distribution function (PDF) of the aperture mass (M_{ap}), a common observable for cosmic shear observations. We show that the LDP can indeed be used in practice for a much larger family of observables than previously envisioned, such as those built from continuous and nonlinear functionals of density profiles. Taking advantage of this formalism, we can extend previous results, which were based on crude definitions of the aperture mass, with top-hat windows and the use of the reduced shear approximation (replacing the reduced shear with the shear itself). We were precisely able to quantify how this latter approximation affects the M_{ap} statistical properties. In particular, we derive the corrective term for the skewness of the M_{ap} and reconstruct its one-point PDF.

DOI: [10.1103/PhysRevD.97.023524](https://doi.org/10.1103/PhysRevD.97.023524)

I. INTRODUCTION

Inferring statistical properties of the cosmic density fields beyond the linear regime is a daunting task that would have far-reaching implications for the scientific exploitations of large-scale surveys now available or under construction. Recently, the large deviation principle (LDP) has been put forward as a general framework for doing such calculations for well-chosen observables [1–4]. It shows why and how leading-order cumulant generating functions of various quantities, such as density in spherical or cylindrical cells, can be computed. It gives incidentally the reason why the tree-order cumulant generating function, as derived in standard references such as [5–7], is intimately related to the spherical (or cylindrical) collapse dynamics.

It is beyond the scope of this paper to review the LDP. Let us simply remind the reader that the LDP [8–10] is a branch of probability theory that deals with the rate at which probabilities of certain events decay as a natural parameter of the problem varies [11], and it is applied in a variety of domains in mathematics and theoretical physics, especially in statistical physics, both for equilibrium and nonequilibrium systems (see, for instance, [12] for a review paper on the subject). The application of the LDP to large-scale-structure cosmology has been formalized in [3] and will be employed here in the context of cosmic shear observations in order to demonstrate the reach of this

approach. The practical implementation of this method will be presented afterwards.

We think this method is particularly suited for such observations. Gravitational lensing has indeed been shown to be a very efficient way of exploring the properties of the mass distribution at a large scale. It 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 arclike 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 evidence that large-scale structure can coherently affect the shape of background galaxies was presented in 2000 in a series of compelling results [13–15]. 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 the 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 a critical region), the information provided by cosmic shear observations is 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 [16–18]. 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 the approach that is usually adopted in projects such as the CFHTLenS survey [19] and the DES survey [20]. There exist, however, alternative approaches that give complementary information. In particular, this is the case for 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 that we define below) and its first few moments as indicated in [5,21,22]. It has been exploited in [23] for the CFHTLS and in [24] for the DES survey.

The investigations carried out in this paper are in this line of investigation. We do not aim, however, to quantify the efficiency of such measurements in constraining cosmological parameters, but rather to show that the one-point PDF of the aperture mass can be computed from first principles in a given cosmological context. This is not the first time that arguments related to the theory of large deviations have been used in this context, and we highlight [5] in particular. In this paper we will reconsider this problem and extend the results that had been obtained in two ways: we define the aperture mass with the help of more realistic filters, and we show that one can take into account the fact that only the reduced shear is accessible to observations, 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 observations 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 the aperture,⁵ that is, the aperture mass (M_{ap}) [25,26].

In general, observations should then be viewed as a nonlocal, nonlinear transformation of the convergence

field. The reduced shear is usually replaced by the shear itself in the literature under the argument that the convergence ought to be generally small for cosmic shear observations. It dramatically simplifies the problem as the observed M_{ap} is made into 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 therefore 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 from the variance. What the LDP can account for is then precisely the impact of excursions of large values of the convergence.

The general techniques for computing one-point PDFs of densities filtered with top-hat filters have been developed in a long series of papers [2,7,27,28]. However, the derivation of exact results for continuous filters escaped 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 [3]. The second main point of this paper is to show that the LDP can be applied to a nonlinear functional of the density profile and, namely, to the one-point PDF of the *reduced* M_{ap} . Again, this construction will make use of the general formulation developed in [3]. Our work therefore extends previous results in these two fundamental ways.

The application of the LDP is based on a number of ingredients that we will detail in Sec. II. The line of reasoning that we follow (similar to [5]) is based on the fact that cosmic shear observations are akin to observations in long cylindrical cells. In the following we will 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:

- (i) One should first define the rate function of the variables that define the initial field configuration. We will assume here a Gaussian initial condition to define their covariance matrix.
- (ii) 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).
- (iii) We are then in position to write the observable—say, the M_{ap} defined with a specific filter—as a functional form of the final and therefore initial mass profile.

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

¹www.cfhtlens.org

²www.darkenergysurvey.org

³www.lsst.org

⁴www.euclid-ec.org

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

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

II. THE LDP APPLIED TO THE REDUCED APERTURE MASS

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 that we are interested in.

A. The convergence

Lensing effects are generally classified as strong and weak. In the first case, arcs and multiple images are observable as a consequence of caustics and cusps on the observer's past-light cone due to gravitation. Weak gravitational lensing is associated with smooth deformations on the observer's past-light cone, and the effects are less dramatic. Gravitational lensing of 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 Friedmann-Lemaître-Robertson-Walker spacetime, convergence and shear can be expressed as derivatives of a lens potential. 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 in the definition of κ to the 3D Laplacian, we can use the Poisson equation to link convergence and matter density fluctuations as (see for instance [17])

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

Here χ_S is the comoving distance to the sources, and

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

where \mathcal{D}_0 is the angular distance. 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

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

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

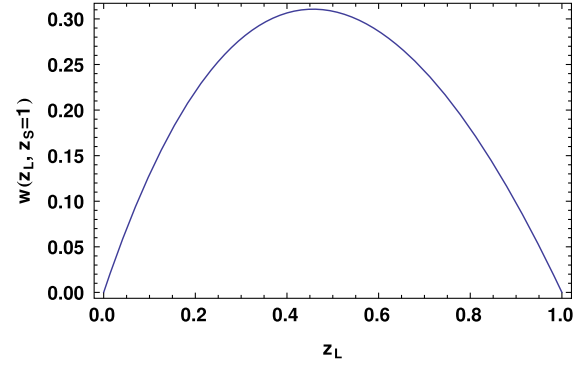


FIG. 1. Projection factor $w(z_L, z_S)$ defined in Eq. (2) in an Einstein–de Sitter universe for $z_S = 1$. The effective value w_{eff} is the particular values of $w(z_L, z_S)$ at $z_L = z_L^*$ corresponding to the redshift of the effective lens plan.

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

This corresponds effectively to the existence of a lens plan at χ_L^* , where the entire lensing mass would be concentrated. We could have defined a projected density contrast instead, as in [5], but the assumption of a lens plan is enough for our present argument. In an Einstein–de Sitter universe, the possible values of $w_{\text{eff}}(z_S)$ are shown in Fig. 1 as a function of the possible z_L^* 's.

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

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

and it plays the role of the projected density contrast. We will also assume a small angle approximation and consider that the integral in Eq. (5) is performed in a cylindrical region instead of a conic one. We refer to [5,28,30] for a detailed discussion of the approximations assumed here.

For reasons that will be explicit in the next paragraph, we are led to consider the smoothed convergence over a cylindrical region with aperture scale θ_i ,

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

for a given sequence of scales $\theta_1, \theta_2, \dots, \theta_N$ 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

$$\delta_{<}(\boldsymbol{\theta}) = \frac{\kappa_{<}(\boldsymbol{\theta})}{w_{\text{eff}}(\chi_S)} \quad (7)$$

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 $\delta_{<}^{\text{lin}}(\theta)$ as in Eq. (7), but using τ_{2D} in Eq. (3). We want to compute the elements $\Sigma_{ij} := \langle \delta_{<}^{\text{lin}}(\theta_i) \delta_{<}^{\text{lin}}(\theta_j) \rangle$, $1 \leq i, j \leq N$, of the covariance matrix Σ for this Gaussian field. In the small-angle approximation, one obtains [28]

$$\Sigma_{ij} = \int_0^\infty \frac{dk_\perp k_\perp}{2\pi} P(k_\perp) W(D_0 \theta_i k_\perp) W(D_0 \theta_j k_\perp), \quad (8)$$

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 the current time. We assume here $P(k_\perp) = Ak_\perp^n$, $-2 \leq n \leq 1$.

As observed in Appendix A, the collection $\{\delta_{<}^{\text{lin}}(\theta_i)\}_{1 \leq i \leq N}$ of correlated Gaussian random variables obeys the LDP with rate function,

$$I(\delta_{<}^{\text{lin}}(\theta_1), \dots, \delta_{<}^{\text{lin}}(\theta_N)) = \frac{\sigma^2(\theta_N)}{2} \sum_{ij} \Xi_{ij} \delta_{<}^{\text{lin}}(\theta_i) \delta_{<}^{\text{lin}}(\theta_j), \quad (10)$$

where $\Xi = \Sigma^{-1}$ and $\sigma^2(\theta_N) = \Sigma_{NN}$. When we take the limit $\sigma^2(\theta_N) \rightarrow 0$, the rate function determines the exponential decay rate for the probability density function associated with 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 noninteracting particles obeying continuity, Euler, and Poisson equations with azimuthal symmetry. If this dynamics can be solved, a map connecting linear and nonlinear overdensities can be established. If we consider matter contained in a cylindrical region, the Gauss theorem will provide us with the relation of the radii given the initial and final density. For this work, it is sufficient to know that the normalized nonlinear density can be approximated in terms of the linear density as [28]

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

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

$$\delta_{<}(\vartheta_i) = \zeta(\delta_{<}^{\text{lin}}(\theta_i)) - 1, \quad \vartheta_i = \theta_i / \zeta(\delta_{<}^{\text{lin}}(\theta_i))^{1/2}. \quad (12)$$

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 $\{\delta_{<}(\vartheta_i)\}_{1 \leq i \leq N}$ is obtained from $\{\delta_{<}^{\text{lin}}(\theta_i)\}_{1 \leq i \leq N}$ by the continuous function ζ , and therefore the new family of random variables also obeys the LDP as a consequence of the contraction principle in large deviations theory (see Appendix A).

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. (A4)]

$$\Psi(\delta_{<}(\vartheta_1), \dots, \delta_{<}(\vartheta_N)) = \inf_{\delta_{<}^{\text{lin}}} I(\delta_{<}^{\text{lin}}(\theta_1), \dots, \delta_{<}^{\text{lin}}(\theta_N)), \quad (13)$$

where $\inf_{\delta_{<}^{\text{lin}}}$ stands for the infimum taken over the collection $\{\delta_{<}^{\text{lin}}(\theta_i)\}_{(1 \leq i \leq N)}$ such that $\delta_{<}(\vartheta_i) = \zeta(\delta_{<}^{\text{lin}}(\theta_i)) - 1$. In the domain in which ζ is bounded, we can perform the inversion $\delta_{<}^{\text{lin}}(\theta_i) = \zeta^{-1}[1 + \delta_{<}(\vartheta_i)]$, so that we may also write $\delta_{<}^{\text{lin}}(\delta_{<}(\vartheta_i))$. We can therefore write

$$\Psi(\delta_{<}(\vartheta_1), \dots, \delta_{<}(\vartheta_N)) = \frac{\sigma^2(\vartheta_N)}{2} \sum_{ij} \Xi_{ij} \delta_{<}^{\text{lin}}(\delta_{<}(\vartheta_i)) \delta_{<}^{\text{lin}}(\delta_{<}(\vartheta_j)). \quad (14)$$

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 A).

E. The single cell case

In order to summarize and illustrate the rate function, SCGF, and their relations and role in 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

$$\Psi(\delta_{<}(\vartheta)) = \frac{\sigma^2(\vartheta) (\delta_{<}^{\text{lin}}(\delta_{<}(\vartheta)))^2}{2\sigma^2(\vartheta \zeta^{1/2}(\delta_{<}^{\text{lin}}(\vartheta)))}. \quad (15)$$

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 $\Psi(\delta_{<})$ shown in Fig. 2 that this function is not globally convex. Indeed, there is a critical value $\delta_{<}^c$ where there is a change of convexity. For $-2 \leq n \leq 1$, however, $\delta_{<}^c > 0$, indicating that the rate function is convex in a neighborhood of the origin.

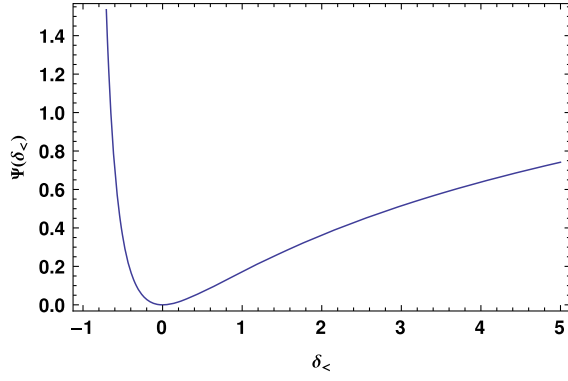


FIG. 2. The rate function given in Eq. (15) for $n = -1.5$. The rate function is convex around the origin but changes the convexity for $\delta_{<} \approx 0.8$.

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

$$\varphi(\lambda) = \sup_{\delta_{<}} [\lambda \delta_{<} - \Psi(\delta_{<})]. \quad (16)$$

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 Ψ . Moreover, $\varphi = \Psi^*$ has points of nondifferentiability when Ψ loses convexity. Our rate function, as observed, is not globally convex for all values of the spectral index n , which requires careful analysis of the inversion of φ to obtain the PDF [2,4].

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

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

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

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

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

III. THE LDP FORMULATION APPLIED TO THE REDUCED 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 with scales $\mathcal{D}_0 \vartheta_1, \dots, \mathcal{D}_0 \vartheta_N$ 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 $\int dx x U(x) = 0$. In order to keep working with normalized quantities, we shall define the normalized aperture mass as in [5]:

$$\delta_{\text{ap}} = U * \delta_{\text{proj}}. \quad (19)$$

Here, $\delta_{\text{proj}}(\vartheta) := \kappa(\vartheta)/w_{\text{eff}}$. 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 $\delta_{<}$ on the definition. We remark that the differentiation of Eq. (6) yields $\delta_{\text{proj}}(\vartheta) = \kappa(\vartheta)/w_{\text{eff}} = \delta_{<}(\vartheta) + \frac{\vartheta}{2} \delta'_{<}(\vartheta)$. We can thus use integration by parts to write

$$\delta_{\text{ap}} = \int d\vartheta \vartheta U(\vartheta) \delta_{\text{proj}}(\vartheta) = \int d\vartheta \tilde{U}(\vartheta) \delta_{<}(\vartheta), \quad (20)$$

where $\tilde{U}(\vartheta) = -\vartheta^2 U'(\vartheta)/2$. The filter \tilde{U} satisfies $\int d\vartheta \tilde{U}(\vartheta) = 0$. Going one step further, we can express the δ_{ap} in terms of a specific filter acting on the shear [26]. For axially symmetrical lenses, we can relate δ_{proj} , $\delta_{<}$, and $\gamma^{\text{eff}} = \gamma/w_{\text{eff}}$ as

$$\gamma^{\text{eff}}(\vartheta) = \delta_{<}(\vartheta) - \delta_{\text{proj}}(\vartheta) = -\frac{\vartheta}{2} \delta'_{<}(\vartheta). \quad (21)$$

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

$$\begin{aligned} \delta_{\text{ap}} &= \int d\vartheta \tilde{Q}(\vartheta) \gamma^{\text{eff}}(\vartheta) \\ &= \int d\vartheta \frac{\tilde{Q}(\vartheta) + \vartheta \tilde{Q}'(\vartheta)}{2} \delta_{<}(\vartheta), \end{aligned} \quad (22)$$

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

$$\tilde{U}(\vartheta) = \frac{1}{2} (\tilde{Q}(\vartheta) + \vartheta \tilde{Q}'(\vartheta)). \quad (23)$$

As before, $\tilde{Q}(\vartheta) = -\vartheta^2 Q'(\vartheta)/2$. For analytical convenience, we will focus on the filter $Q(\vartheta) = e^{-\vartheta^2/2}$. Other possible compensated filters can be found in the literature [5,26,31].

B. Conveniences and limitations of the aperture mass

The convergence field can only be obtained through inversion and is nonlocal. Shear field is more directly observed. Observations of ellipticity fields, however, yield measures of the reduced shear, from which the shear is usually obtained by assuming $\kappa \ll 1$ for weak lensing [17,18]. This leads us to define

$$\delta_{\text{ap}}^g = \int d\vartheta g^{\text{eff}}(\vartheta) \tilde{Q}(\vartheta). \quad (24)$$

One should then distinguish δ_{ap}^g and δ_{ap} in Eq. (22), which agree only as long as the convergence is small. Being defined as the filtering of a random field, however, the convergence can exhibit statistically large fluctuations, and therefore the statement $\kappa \ll 1$ is not well defined by itself.

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

C. General filter and the δ_{ap} statistics

We have presented in Sec. II C the formalism based on the large deviations theory to derive the SCGF associated with 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 [3], and we will develop this generalization to produce the SCGF for the δ_{ap} and δ_{ap}^g .

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

$$\delta_{\text{ap}} \approx \sum_{i=1}^N \tilde{U}_i(\vartheta_i) \delta_{<}(\vartheta_i). \quad (25)$$

If this is the case, the δ_{ap} is a linear combination of the $\delta_{<}(\vartheta_i)$'s and therefore, by the contraction principle, the rate function for the δ_{ap} is given in terms of the initial rate function $I(\theta_1, \dots, \theta_N)$ through a composition of continuous bounded functions. By the contraction principle [see Eq. (A4)],

$$\begin{aligned} \Psi(\delta_{\text{ap}}) = & \inf_{\delta_{<}^{\text{lin}}(\theta_i)} \left[I(\delta_{<}^{\text{lin}}(\theta_1), \dots, \delta_{<}^{\text{lin}}(\theta_N)) \right. \\ & \left. + \alpha \left(\delta_{\text{ap}} - \sum_{i=1}^N \tilde{U}_i(\vartheta_i) \delta_{<}(\vartheta_i) \right) \right], \quad (26) \end{aligned}$$

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

The weights \tilde{U}_i are defined on the image of ζ , i.e., on the filtered nonlinear field, but the inf has to be computed over the initial random variables $\delta_{<}^{\text{lin}}(\theta_1), \dots, \delta_{<}^{\text{lin}}(\theta_N)$, both being related by the map ζ defined in Eqs. (11) and (12).

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

$$\begin{aligned} \sum_{i=1}^N \tilde{U}_i(\vartheta_i) \delta_{<}(\vartheta_i) & \rightarrow \int d\vartheta \tilde{U}(\vartheta) \zeta(\delta_{<}^{\text{lin}}(\theta)) \\ & = \int d\theta \frac{d\vartheta_{\delta_{<}^{\text{lin}}}}{d\theta} \tilde{U}[\vartheta_{\delta_{<}^{\text{lin}}}(\theta)] \zeta(\delta_{<}^{\text{lin}}(\theta)), \quad (27) \end{aligned}$$

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

$$\vartheta_{\delta_{<}^{\text{lin}}}(\theta) = \theta \zeta^{-1/2}(\delta_{<}^{\text{lin}}(\theta)). \quad (28)$$

D. The scaled cumulant generating functions of δ_{ap}

Our next goal is write the SCGF on the continuous limit. By the Varadhan lemma given in Eq. [8], the scaled cumulant generating function will be the continuous limit of

$$\varphi(\lambda) = \sup_{\delta_{<}^{\text{lin}}} \left[\lambda \sum_{i=1}^N \tilde{U}_i(\vartheta_i) \delta_{<}(\vartheta_i) - I(\delta_{<}^{\text{lin}}(\theta_1), \dots, \delta_{<}^{\text{lin}}(\theta_N)) \right]. \quad (29)$$

In order to write the continuous limit of the rate function $I(\delta_{<}^{\text{lin}}(\theta_1), \dots, \delta_{<}^{\text{lin}}(\theta_N))$ 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

$$\int d\theta' \sigma^2(\theta, \theta') \xi(\theta', \theta'') = \delta_{\text{D}}(\theta - \theta''), \quad (30)$$

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

The desired continuous limit for the SCGF is, therefore,

$$\begin{aligned} \varphi(\lambda) = & \sup_{\delta_{<}^{\text{lin}}} \left[\lambda \int d\theta \tilde{U}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta)) \frac{d\vartheta_{\delta_{<}^{\text{lin}}}}{d\theta} \zeta(\delta_{<}^{\text{lin}}(\theta)) \right. \\ & \left. - \frac{\sigma_F^2}{2} \int d\theta d\theta' \delta_{<}^{\text{lin}}(\theta) \delta_{<}^{\text{lin}}(\theta') \xi(\theta, \theta') \right]. \quad (31) \end{aligned}$$

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

As already observed, the normalized aperture mass δ_{ap} can be obtained as in Eq. (20) through the convolution of

$\delta_{<}$ and \tilde{U} , or as in Eq. (22) by convolving γ^{eff} and \tilde{Q} , and the two different expressions are related by an integration by parts. Integrations by parts can also be applied to Eq. (31) to reexpress it explicitly in terms of the convergence or shear with corresponding filters. Although equivalent, the different expressions of φ allow us to perform distinct numerical implementations and check the performance of the solutions. A second need for a rewriting of Eq. (31) in terms of γ^{eff} and \tilde{Q} is to extend the SCGF for δ_{ap} into the SCGF of δ_{ap}^g . We will call the different arrangement of variables “representations” of the SCGF.

1. SCGF on the convergence representation

The SCGF for δ_{ap} in Eq. (31) is already given in terms of $\delta_{<}$ and \tilde{U} , but it can be expressed in a more suitable form. Let the filter $\tilde{V}(x)$ be defined by

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

It follows by integration by parts that

$$\begin{aligned} & \int_0^\infty d\theta \left(\tilde{U}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta)) \frac{d\vartheta_{\delta_{<}^{\text{lin}}}}{d\theta} \right) \zeta(\delta_{<}^{\text{lin}}(\theta)) \\ &= - \int_0^\infty \tilde{V}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta)) \frac{\partial \zeta(\delta_{<}^{\text{lin}}(\theta))}{\partial \delta_{<}^{\text{lin}}} \delta_{<}^{\text{lin}'}(\theta), \end{aligned} \quad (33)$$

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

$$\begin{aligned} \varphi^\kappa(\lambda) &= -\inf_{\delta_{<}^{\text{lin}}} \left[\lambda \int d\theta \tilde{V}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta')) \frac{\partial \zeta(\delta_{<}^{\text{lin}}(\theta'))}{\partial \delta_{<}^{\text{lin}}} \delta_{<}^{\text{lin}'}(\theta') \right. \\ & \left. + \frac{\sigma_F^2}{2} \int d\theta d\theta' \delta_{<}^{\text{lin}}(\theta) \delta_{<}^{\text{lin}}(\theta') \xi(\theta, \theta') \right]. \end{aligned} \quad (34)$$

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

$$\begin{aligned} \varphi^\gamma(\lambda) &= -\inf_{\delta_{<}^{\text{lin}}} \left[\lambda \int d\theta \tilde{Q}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta)) \frac{\vartheta_{\delta_{<}^{\text{lin}}}}{2} \frac{d\zeta(\delta_{<}^{\text{lin}}(\theta))}{d\theta} \right. \\ & \left. + \frac{\sigma_F^2}{2} \int d\theta d\theta' \delta_{<}^{\text{lin}}(\theta) \delta_{<}^{\text{lin}}(\theta') \xi(\theta, \theta') \right]. \end{aligned} \quad (35)$$

We recall that $\gamma^{\text{eff}} = -\frac{\vartheta_{\delta_{<}^{\text{lin}}}}{2} \frac{d\zeta(\delta_{<}^{\text{lin}}(\theta))}{d\vartheta_{\delta_{<}^{\text{lin}}}}$.

It should be stressed that in the continuous limit we have

$$\varphi^\kappa(\lambda) \equiv \varphi^\gamma(\lambda). \quad (36)$$

The two formulations will depart, however, once the integral is computed with discrete points. It will be used in the following as a convergence test of the numerical resolution of the infimum problem.

E. The scaled cumulant generating function of δ_{ap}^g , the physical and observable M_{ap}

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

$$\begin{aligned} g^{\text{eff}}(\delta_{<}^{\text{lin}}(\theta)) &= \frac{\gamma^{\text{eff}}}{1 - \kappa} \\ &= \frac{-(\vartheta/2)\delta_{<}'(\vartheta)}{1 - w_{\text{eff}}(\chi_s)(\delta_{<}(\vartheta) - \frac{1}{2}\vartheta\delta_{<}'(\vartheta)')} \end{aligned} \quad (37)$$

in terms of the variables ϑ , $\delta_{<}(\vartheta)$, and the projection factor w_{eff} . The contraction principle allows us to extend the LDP to any bounded continuous functional of the initial profile. We can invoke it to replace γ^{eff} by g^{eff} in Eq. (35) and obtain the SCGF on the reduced shear representation:

$$\begin{aligned} \varphi^g(\lambda) &= -\inf_{\delta_{<}^{\text{lin}}} \left[\lambda \int d\theta \tilde{Q}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta)) g^{\text{eff}}(\delta_{<}^{\text{lin}}(\theta)) \right. \\ & \left. + \frac{\sigma_F^2}{2} \int d\theta d\theta' \delta_{<}^{\text{lin}}(\theta) \delta_{<}^{\text{lin}}(\theta') \xi(\theta, \theta') \right]. \end{aligned} \quad (38)$$

This equation generalizes Eq. (35) for the reduced shear [there is, however, no alternative representation of $\varphi^g(\lambda)$ as for δ_{ap}]. In terms of mathematical construction, $\varphi^g(\lambda)$ is an example of the explicit expression of a SCGF for a continuous nonlinear functional form of the initial density profile encoded in $\delta_{<}^{\text{lin}}(\theta)$. It therefore generalizes previous approaches based on cumulant resummations of discrete cells in a dramatic way.

The practical feasibility of the method is, however, based on our ability to actually find the infs that appear in Eqs. (34), (35), and (38). If it can be done, the statistical properties of δ_{ap} and δ_{ap}^g will follow. This is the purpose of the next section.

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 those minimization problems and check that we have consistent results. Details on the numerical tests can be found in Appendix B.

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

The problem we are facing corresponds to the minimization of an action and it can be looked for by imposing that the variation of the action vanish at the desired solution. Formally, an extremum of the functional $\varphi^\#(\lambda)$ for any of the SCGF we considered can be obtained by imposing

$$\frac{\delta\varphi^\#(\lambda)}{\delta\delta_{<}^{\text{lin}}(\theta)} = 0. \quad (39)$$

Such a constraint will lead to the writing of the Euler-Lagrange equations for the initial density profile.

Its implementation is quite involved in general but can be easily done for $\varphi^\kappa(\lambda)$ as it is a simple function of $\delta_{<}^{\text{lin}}(\theta)$. Indeed, it follows from Eqs. (39), (34), and (33) that the linear profile has to obey the equation

$$\begin{aligned} \delta_{<}^{\text{lin}}(\theta) = & -\frac{\lambda}{\sigma_F^2} \int d\theta' \sigma^2(\theta, \theta') \frac{\delta}{\delta\delta_{<}^{\text{lin}}} \left[\tilde{V}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta')) \right. \\ & \left. \times \frac{\partial\zeta(\delta_{<}^{\text{lin}}(\theta'))}{\partial\delta_{<}^{\text{lin}}} \delta_{<}^{\text{lin}}(\theta') \right], \end{aligned} \quad (40)$$

where

$$\frac{\delta}{\delta\delta_{<}^{\text{lin}}} = \frac{\partial}{\partial\delta_{<}^{\text{lin}}} - \frac{d}{d\theta} \frac{\partial}{\partial\delta_{<}^{\text{lin}}} \quad (41)$$

is the Euler-Lagrange operator. The term inside squared brackets in Eq. (40) has the nice property of being linear on $\delta_{<}^{\text{lin}}$, which leads to great simplifications. Indeed, the use of the explicit form of $\vartheta_{\delta_{<}^{\text{lin}}}$ given in Eq. (28) and the definition of \tilde{V} given in Eq. (32) lead directly to

$$\begin{aligned} \delta_{<}^{\text{lin}}(\theta) = & \frac{\lambda}{\sigma_F^2} \int d\theta' \sigma^2(\theta, \theta') \left[\tilde{U}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta')) \right. \\ & \left. \times \zeta^{-1/2}(\delta_{<}^{\text{lin}}(\theta')) \frac{\partial\zeta(\delta_{<}^{\text{lin}}(\theta'))}{\partial\delta_{<}^{\text{lin}}} \right]. \end{aligned} \quad (42)$$

Also from Eq. (39) it follows that

$$\begin{aligned} & \frac{\sigma_F^2}{2} \int d\theta d\theta' \delta_{<}^{\text{lin}}(\theta) \delta_{<}^{\text{lin}}(\theta') \xi(\theta, \theta') \\ & = \frac{\lambda}{2} \int d\theta d\theta' \delta_{<}^{\text{lin}}(\theta) \\ & \quad \times \left[\tilde{U}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta')) \zeta^{-1/2}(\delta_{<}^{\text{lin}}(\theta')) \frac{\partial\zeta(\delta_{<}^{\text{lin}}(\theta'))}{\partial\delta_{<}^{\text{lin}}} \right], \end{aligned} \quad (43)$$

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,

$$\begin{aligned} \varphi^\kappa(\lambda) = & -\lambda \int d\theta \tilde{V}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta; \lambda)) \frac{\partial\zeta(\delta_{<}^{\text{lin}}(\theta; \lambda))}{\partial\delta_{<}^{\text{lin}}} \delta_{<}^{\text{lin}}(\theta; \lambda) \\ & - \frac{\lambda}{2} \int d\theta d\theta' \delta_{<}^{\text{lin}}(\theta; \lambda) \left[\tilde{U}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta'; \lambda)) \right. \\ & \left. \times \zeta^{-1/2}(\delta_{<}^{\text{lin}}(\theta'; \lambda)) \frac{\partial\zeta(\delta_{<}^{\text{lin}}(\theta'; \lambda))}{\partial\delta_{<}^{\text{lin}}} \right], \end{aligned} \quad (44)$$

where $\delta_{<}^{\text{lin}}(\theta; \lambda)$ is a solution of the integral equation (42) for each value of λ , and $\vartheta_{\delta_{<}^{\text{lin}}}(\theta; \lambda) = \theta \zeta^{-1/2}(\delta_{<}^{\text{lin}}(\theta; \lambda))$.

The linearity of our action⁷ in $\delta_{<}^{\text{lin}}$ allows us to obtain the simple expression for $\delta_{<}^{\text{lin}}(\theta)$ given in Eq. (42). We will assume that $\delta_{<}^{\text{lin}}(\theta)$ conducts to a maximum of the SCGF on the interval on which it is defined.

The SCGF is given by Eq. (44) as long as Eq. (42) can be solved for each value of λ . In order to study the existence of unique solutions for this equation, we take the first variation of Eq. (42), obtaining

$$\begin{aligned} & \frac{\sigma^2(\theta, \theta')}{\sigma_F^2} \frac{\delta}{\delta\delta_{<}^{\text{lin}}(\theta)} \left[\tilde{U}(\vartheta_{\delta_{<}^{\text{lin}}}(\theta')) \zeta^{-1/2}(\delta_{<}^{\text{lin}}(\theta')) \frac{\partial\zeta(\delta_{<}^{\text{lin}}(\theta'))}{\partial\delta_{<}^{\text{lin}}} \right] \\ & = \frac{\delta_D(\theta - \theta')}{\lambda}, \end{aligned} \quad (45)$$

from which we can determine critical values of λ on the solution of Eq. (42).

B. Skewness of δ_{ap} and δ_{ap}^g

Expanding $\varphi(\lambda)$ given in Eq. (31) around $\lambda = 0$ leads to the expression of cumulants. For the skewness we have [3]

$$\hat{s}_3^\kappa = 3\nu_2 \frac{\int dx \tilde{U}(x) \Sigma^2(x)}{[\int dx \tilde{U}(x) \Sigma(x)]^2} + 3 \frac{\int dx x \tilde{U}(x) \Sigma(x) \Sigma'(x)}{[\int dx \tilde{U}(x) \Sigma(x)]^2}, \quad (46)$$

with

$$\Sigma(x) = \int dy \sigma^2(x, y) \tilde{U}(y). \quad (47)$$

[Note that the coefficient in front of the second term in Eq. (46) is $6/d$, where d is the dimension of space in which the spherical collapse is considered.] It is worth recalling that the reduced skewness of the aperture mass is then

⁷On the other side, the same linearity of the action in $\delta_{<}^{\text{lin}}$ forbids us from exploring the stability of the extremum through second variations of the action.

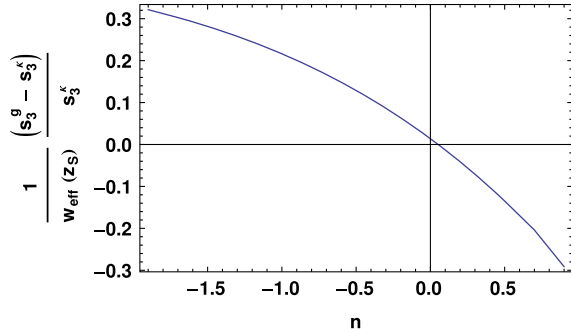


FIG. 3. Deviation of the skewness for δ_{ap}^g and δ_{ap} normalized by w_{eff} . Since $0 \leq w_{\text{eff}} \leq 0.3$ in Einstein–de Sitter for $z_S = 1$, we see that the highest possible correction to the skewness in this case is of the order of 10%.

$$s_3^k = \frac{1}{w_{\text{eff}}} \hat{s}_3^k. \quad (48)$$

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

$$\hat{s}_3^g - \hat{s}_3^k = 6w_{\text{eff}} \frac{\int dx \tilde{Q}(x) \frac{x}{2} \Sigma'(x) (\Sigma(x) + \frac{x}{2} \Sigma'(x))}{[\int dx \tilde{U}(x) \Sigma(x)]^2} \quad (49)$$

and

$$s_3^g - s_3^k = \frac{1}{w_{\text{eff}}} (\hat{s}_3^g - \hat{s}_3^k). \quad (50)$$

As expected, the expressions of s_3^k and s_3^g depend on the choice of power spectrum and, for the sake of simplicity, we illustrate our results for power law spectra. More specifically, we can analyze the relative importance of the correction term given in Eq. (49) by computing the ratio $(s_3^g - s_3^k)/(w_{\text{eff}} s_3)$, which is then independent of w_{eff} . As shown in Fig. 3, the impact on the skewness of the distribution of δ_{ap}^g has a few percent deviation from the skewness for δ_{ap} as a function of the spectral index n .

As observed in Eq. (18), the p th cumulant is a homogeneous function of degree p , and in particular $s_3 = \hat{s}_3/w_{\text{eff}}$.

C. The SCGF and the rate functions

The Euler-Lagrange equations presented in Sec. IV A can be used to obtain φ^k , but the action becomes considerably more involved when we are interested in φ^r . We treated the problem numerically in two alternative ways: the implementation of Eqs. (42) and (44), or the direct extremization methods available in Mathematica such as the FindMinimum routine. As shown in Appendix B, direct extremization and the Euler-Lagrange methods agree to

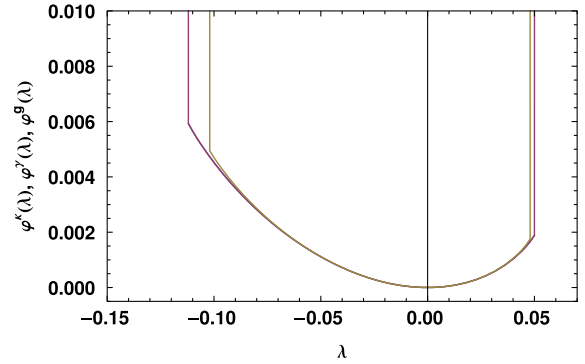


FIG. 4. The SCGFs $\varphi^k(\lambda)$, $\varphi^r(\lambda)$, and $\varphi^g(\lambda)$ for $n = -1.5$. The curves for $\varphi^k(\lambda)$ and $\varphi^r(\lambda)$ coincide as they should. The SCGF for δ_{ap}^g is, however, slightly different and has different critical points. We use $w_{\text{eff}}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

high precision on the range of λ of our interest and for the discretizations of the interval considered.

The general output from the numerical solutions can be seen in Fig. 4. In this plot, $\varphi^k(\lambda)$ is computed using the Euler-Lagrange strategy and $\varphi^r(\lambda)$ and $\varphi^g(\lambda)$ are derived from direct extremization. We should remark that although they are displayed in two different representations, we have $\varphi^k(\lambda) = \varphi^r(\lambda)$, and this is recovered in the numerical reconstructions (see Appendix B 1 for more details). We note that moving from the shear to the reduced shear induces some changes in the SCGF and also moves the critical points closer to the origin. The critical points for φ^k can also be obtained from Eq. (45).

The SCGF is always a convex function; the same is not true for the rate function. As already noted, and discussed in Appendix A, the SCGF and rate function can be obtained from each other by a 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 it will present points of nondifferentiability. 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.⁸

⁸A duality property connects features of the SCGF and the convex envelope of the rate function [12]: 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 imply 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 nonaffine segment of the convex envelope just described.

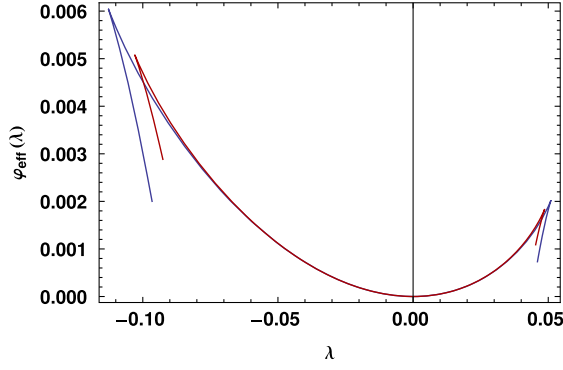


FIG. 5. The effective cumulant generating functions $\varphi_{\text{eff}}^{\kappa}$ and φ_{eff}^g satisfying Eq. (51). The projection factor $w_{\text{eff}} = 0.1$ is used on the φ_{eff}^g data.

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 δ_{ap} values. The calculation is based on the computation of the inverse Laplace transform of the cumulant generating function (see, for instance, [2] for details). For the sake of simplicity, we perform this calculation using an effective form for the SCGF (as in [5]). Such an effective form is based on an effective vertex generating function $\zeta_{\text{eff}}(\delta_{<}^{\text{lin}})$, satisfying

$$\varphi_{\text{eff}}(\lambda) = \lambda \zeta_{\text{eff}}(\delta_{<}^{\text{lin}}) - \frac{1}{2} \lambda \delta_{<}^{\text{lin}} \zeta'_{\text{eff}}(\delta_{<}^{\text{lin}}) \quad (51a)$$

$$\delta_{<}^{\text{lin}} = \lambda \zeta'_{\text{eff}}(\delta_{<}^{\text{lin}}), \quad (51b)$$

where $\zeta_{\text{eff}}(\delta_{<}^{\text{lin}})$ is adjusted so that $\varphi_{\text{eff}}(\lambda)$ provides a good fit to the SCGF we computed, in particular reproducing its critical behaviors as shown in Fig. 5. In practice one can get a very good fit with a fifth-order polynomial for ζ_{eff} .

The effective cumulant generating function obtained this way reproduces extremely accurately the global behavior obtained previously, in particular for its critical points.

The reconstruction of the one-point PDF of δ_{ap} is then obtained from the following form,

$$P(\delta_{\text{ap}}) = \int_{-\infty}^{\infty} \frac{d\lambda}{2\pi} \exp[-\lambda \delta_{<} + \varphi_{\delta_{<}}(\lambda)], \quad (52)$$

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

$$\varphi_{\delta_{<}}(\lambda) = \frac{1}{\delta^2} \varphi^{\kappa}(\lambda \delta^2), \quad (53)$$

in such a way that δ^2 matches the expected variance of δ_{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.

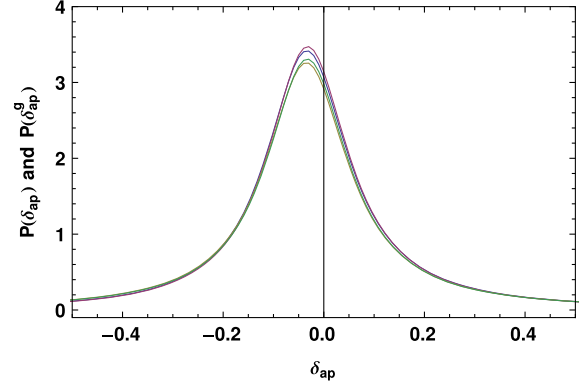


FIG. 6. PDFs obtained from the inverse Laplace transformation of φ^{κ} and φ^g for $\hat{\sigma} = 0.4$ (top two curves) and $\hat{\sigma} = 0.7$ (bottom two curves). The δ_{ap}^g is reconstructed for $w_{\text{eff}} = 0.1$. In each case, $P(\delta_{\text{ap}}^g)$ is slightly larger than $P(\delta_{\text{ap}})$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

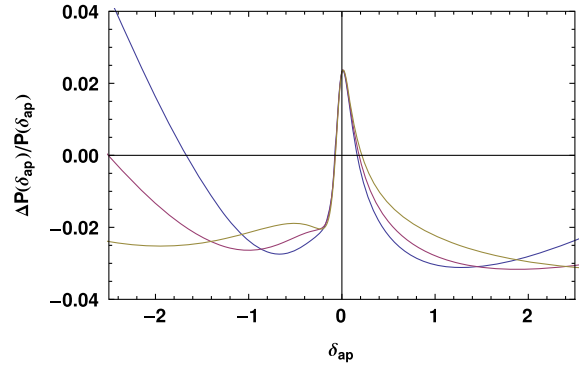


FIG. 7. Difference of the PDFs shown in Fig. 6 for $\hat{\sigma} = 0.4$ (blue curve), $\hat{\sigma} = 0.5$ (red curve), and $\hat{\sigma} = 0.7$ (yellow curve).

The resulting PDFs, $P(\delta_{\text{ap}})$ and $P(\delta_{\text{ap}}^g)$, built, respectively, from the shear field and the reduced shear field, are shown in Fig. 6. For these ranges of δ_{ap} and this value of w_{eff} , the relative errors are about a few percent as shown in Fig. 7, consistent with our findings concerning the skewness. What these results show, however, is that the extra nonlinearities contained in the reduced shear expression have little effect on the global shape of the PDFs.

V. CONCLUSION

We have demonstrated in this study that it is possible to take advantage of the LDP to compute scaled cumulant generating functions and the corresponding PDF for a wide range of observables, namely, the derivation of statistical properties of quantities obtained by general (symmetric) filtering of functionals of the density profiles. The derivation of Eq. (38) is an example of such a construction. It gives there the scaled cumulant generating function of a nonlinear functional of the density profile obtained through a minimization problem. We subsequently showed that it

can be successfully implemented in practice. In particular we have demonstrated in Sec. IV that the numerical results that we obtained were solid in different ways: by using the solution of the Euler-Lagrange equation of the minimization problem when possible and by comparing our results at the level of the third-order cumulant, which can be computed explicitly.

It gives us the opportunity to derive the cumulant generating function and the corresponding one-point PDF of the aperture mass, δ_{ap}^g , 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 fully 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. Such exploration could only be done in the context of the LDP and is unattainable from direct cumulant resummations, as were previously used (as pioneered in [6,7] and extended in [2], for instance).

Formally, the construction of the scaled cumulant generating function is based on the minimization of the 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 detail how to do such a minimization efficiently but checked that our results were correct from different approaches (that are equivalent in the continuous limit),⁹ leaving room for further improvements.

What we have found is that the results that were previously obtained, specifically in [5], 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. We find, however, that the general properties of the cumulant generating functions are left unchanged: for instance, they exhibit critical values for both positive and negative values of the parameters whether the δ_{ap} is built from the shear or 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 nonlinear effects. The impact of this effect is, however, rather mild. We have expressed it in terms of the δ_{ap} skewness and the δ_{ap} one-point PDF.

⁹We note that finding a profile satisfying a given condition can be reexpressed 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.

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

ACKNOWLEDGMENTS

This work was supported by Grant No. ANR-12-BS05-0002 and Labex ILP (reference ANR-10-LABX-63) part of the Idex SUPER of the program Investissements d'avenir Grant No. ANR-11-IDEX-0004-02.

APPENDIX A: 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 $\Pr[\lim_{n \rightarrow \infty} X_n = X] = 1$. Alternatively, a sequence is said to “converge in probability” to a given element X if $\lim_{n \rightarrow \infty} \Pr[|X_n - X| \geq \varepsilon] = 0$, 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

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

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 , $\lim_{n \rightarrow \infty} \phi_n(t) = \phi(t)$. The portmanteau theorem in probability theory [32] states that if α_n converges weakly to α on \mathbb{R} , and $C \in \mathbb{R}$ is a closed set, then

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

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

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

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

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

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 a large deviation principle builds on the theorem just mentioned to include the idea that the probability measures associated with rare events are exponentially suppressed, and it introduces the rate function that modulates the exponential decay.

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

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

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

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

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

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

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log P_\varepsilon(A) = -\inf_{x \in A} I(x). \quad (\text{A1})$$

The parameter ε has to be identified as some limiting parameter on each problem of interest. For collections of identical, identically distributed (i.i.d.) random variables, 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 $\lim_{\varepsilon \rightarrow 0} \varepsilon \log P_\varepsilon(\{X_i\} \in A) = -\inf_{x \in A} I(x)$, where we concentrate on Eq. (A1) 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 [10]: 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 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 $x > \mathbb{E}[X_1]$, $\lambda > 0$,

$$\begin{aligned} \Pr(S_k/k \in [x, \infty)) &= \Pr(\lambda S_k \geq k\lambda x) \\ &\leq \exp[-k\lambda x] \mathbb{E}[e^{\lambda S_k}] \\ &= \exp[-k\lambda x] \prod_{i=1}^k \mathbb{E}[e^{\lambda X_i}] \\ &= \exp[-k\lambda x] (\mathbb{E}[e^{\lambda X_1}])^k \\ &= \exp[-k\lambda x + k \log \mathbb{E}[e^{\lambda X_1}]] \\ &= \exp[-k(\lambda x - c(\lambda))], \end{aligned} \quad (\text{A2})$$

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 the rate function and cumulant generating function are related by the 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

$$I(x) = \frac{1}{2} \sum_{ij} (x - \mu)_i \Xi_{ij} (x - \mu)_j, \quad (\text{A3})$$

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

Going back to the second question, we will list two consequences of the LDP: the contraction principle and 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:

$$J(y) = \inf_{x: F(x)=y} I(x). \quad (\text{A4})$$

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" [9]. 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 prohibitive, and therefore the contraction principle is a powerful tool.

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

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left(\frac{F(X^\varepsilon)}{\varepsilon} \right) \right] = \sup_{x \in \mathbb{R}^n} [F(x) - I(x)]. \quad (\text{A5})$$

To have an idea of why this is so, we should remember that, because of the LDP, $\mathbb{E}[\exp(\frac{F(X^\varepsilon)}{\varepsilon})] \approx \exp[\sup_x [F(x) - I(x)]/\varepsilon]$, where we use the Laplace method to give an approximative answer to the integral. At the limit $\varepsilon \rightarrow 0$ 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

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left(\frac{\sum_i \lambda_i X_i^\varepsilon}{\varepsilon} \right) \right] &=: \varphi(\lambda) \\ &= \sup_{x \in \mathbb{R}^n} [\lambda x - I(x)], \end{aligned} \quad (\text{A6})$$

which is the SCGF. All the (scaled) cumulants can be obtained from $\varphi(\lambda)$ by partial differentiation.

Equation (A6) also states that the SCGF and the rate function are convex conjugates of each other or that they 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 can the SCGF 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 be obtained from the SCGF. Indeed, points of nondifferentiability of the SCGF will be associated with points where the rate function loses convexity.

APPENDIX B: NUMERICAL EVALUATIONS

Our search for the scaled cumulant generating functions for the aperture mass and physical aperture mass brought us to φ^κ , φ^γ , and φ^θ , given in Eqs. (34), (35), and (38), respectively. We must now obtain solutions for the minimization problems. The solution is obtained in terms of a vector value $\delta_{<}^{\text{lin}}(\theta_i)$ for regularly spaced values¹⁰ of θ_i . We typically use 30 points; convergence of the results has been tested against 100 points' results.

One possible strategy to approach the solution of the minimization problem is presented in Sec. IV A. The relative simplicity of the functional brings us to $\varphi^\kappa(\lambda)$ in Eq. (44), once we solve Eq. (42) for each value of λ in a range of interest. We will refer to it as the ‘‘Euler-Lagrange (EL) strategy.’’

A second possible path is the direct search for minima using the numerical algorithms such as FindMinimum in Mathematica [33], that we will call the ‘‘direct extremization (DE) strategy.’’

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

We should also remark that the nonlinear 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 nonlinear density:

$$\zeta^{\text{reg}}(\tau_{2D}) = \frac{(1 + \epsilon)^{\nu/2}}{(1 - (\frac{\tau_{2D}}{\nu})^2 + \epsilon)^{\nu/2}}, \quad (\text{B1})$$

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 by comparing the performances of the Euler-Lagrange and direct extremization strategies for the reconstruction of the same object $\varphi^\kappa(\lambda)$.

¹⁰Note that θ_i are radii in the initial space configuration, not in the final configuration. The problem we solve is then in essence different from the multiple radii problem of the standard approach as described in [2].

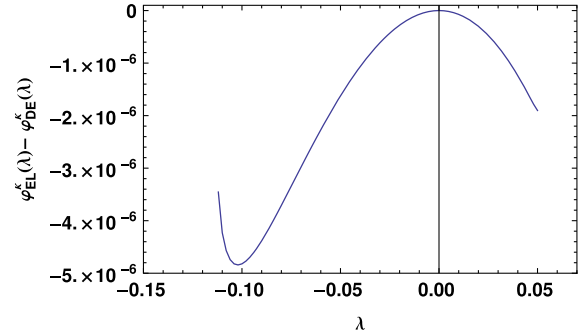


FIG. 8. $\varphi_{EL}^\kappa(\lambda) - \varphi_{DE}^\kappa(\lambda)$ for $n = -1.5$.

a. The SCGF

In order to compare the two numerical strategies for the solution of Eq. (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 validates the DE strategy.

b. Skewness

Given the numerical solution $\varphi_{EL}^\kappa(\lambda)$, we can compute numerically the skewness and compare with the theoretical prediction given in Eq. (46) as in Fig. 9.

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

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%–10% error.

2. The skewness for δ_{ap}^ξ

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 $\varphi^\theta(\lambda)$. In this case, the functional is no longer linear on $\delta_{>}^{\text{lin}}$, and the simplifications obtained in Sec. IV A have

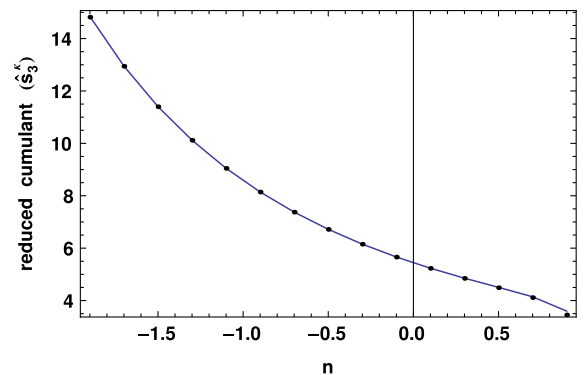


FIG. 9. Theoretical prediction from Eq. (46) and numerical values for the reduced skewness as a function of the spectral index.

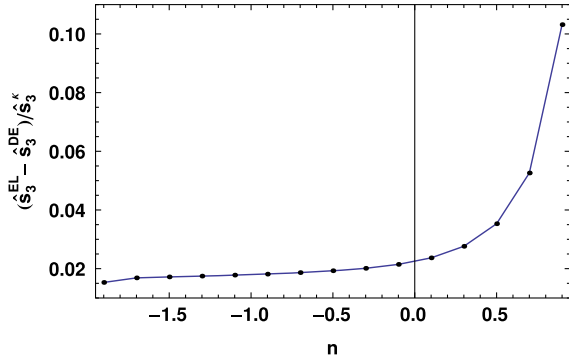


FIG. 10. The difference of the skewness obtained in the EL and DE strategies divided by the theoretical prediction given in Eq. (46). We see that the DE tends to give systematically smaller values for the cumulants for higher values of the spectral index.

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. (49). For this, we compute $\hat{s}_3^g - \hat{s}_3^{\kappa}$ using the DE strategy for $w_{\text{eff}} = 0.1$. The result

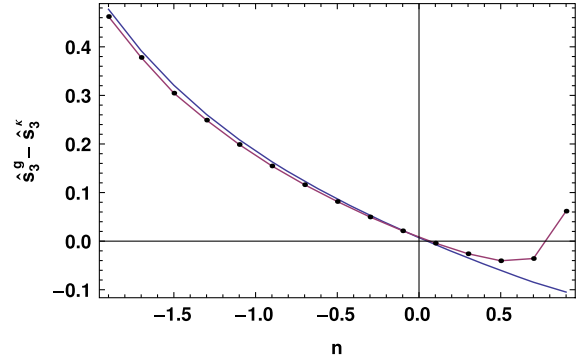


FIG. 11. Theoretical prediction from Eq. (49) and numerical values for the difference of reduced skewness $\hat{s}_3^g - \hat{s}_3^{\kappa}$ as a function of the spectral index for $w_{\text{eff}} = 0.1$.

is shown in Fig. 11. As discussed for Fig. 10, the DE strategy induces 1%–10% error on the skewness, mainly for higher values of n , which also impacts the comparison between numerical points and theoretical prediction in Fig. 11.

-
- [1] P. Valageas, *Astron. Astrophys.* **382**, 412 (2002).
[2] F. Bernardeau, C. Pichon, and S. Codis, *Phys. Rev. D* **90**, 103519 (2014).
[3] F. Bernardeau and P. Reimberg, *Phys. Rev. D* **94**, 063520 (2016).
[4] C. Uhlemann, S. Codis, C. Pichon, F. Bernardeau, and P. Reimberg, *Mon. Not. R. Astron. Soc.* **460**, 1529 (2016).
[5] F. Bernardeau and P. Valageas, *Astron. Astrophys.* **364**, 1 (2000).
[6] F. Bernardeau, *Astrophys. J.* **392**, 1 (1992).
[7] F. Bernardeau, *Astron. Astrophys.* **291**, 697 (1994).
[8] S. R. S. Varadhan, *Large Deviations and Applications* (SIAM, New York, 1984).
[9] F. den Hollander, *Large Deviations* (AMS, Providence, 2008).
[10] R. S. Ellis, Long-Range Interacting Systems, in *The Theory of Large Deviations and Applications to Statistical Mechanics*, Lecture Notes of the Les Houches Summer School Vol. 90, edited by T. Dauxois, S. Ruffo, and L. F. Cugliandolo (Oxford University Press, Oxford, 2008).
[11] S. R. S. Varadhan, Large deviations (2010), <https://math.nyu.edu/faculty/varadhan/LDP/1-2.pdf>.
[12] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
[13] D. M. Wittman, J. A. Tyson, D. Kirkman, I. Dell’Antonio, and G. Bernstein, *Nature (London)* **405**, 143 (2000).
[14] L. Van Waerbeke *et al.*, *Astron. Astrophys.* **358**, 30 (2000).
[15] D. J. Bacon, A. R. Refregier, and R. S. Ellis, *Mon. Not. R. Astron. Soc.* **318**, 625 (2000).
[16] L. Van Waerbeke and Y. Mellier, [arXiv:astro-ph/0305089](https://arxiv.org/abs/astro-ph/0305089).
[17] D. Munshi, P. Valageas, L. Van Waerbeke, and A. Heavens, *Phys. Rep.* **462**, 67 (2008).
[18] M. Kilbinger, *Rep. Prog. Phys.* **78**, 086901 (2015).
[19] L. Fu *et al.*, *Astron. Astrophys.* **479**, 9 (2008).
[20] M. R. Becker *et al.*, *Phys. Rev. D* **94**, 022002 (2016).
[21] D. Munshi and B. Jain, *Mon. Not. R. Astron. Soc.* **318**, 109 (2000).
[22] D. Munshi and B. Jain, *Mon. Not. R. Astron. Soc.* **322**, 107 (2001).
[23] L. Fu *et al.*, *Mon. Not. R. Astron. Soc.* **441**, 2725 (2014).
[24] T. Kacprzak *et al.*, *Mon. Not. R. Astron. Soc.* **463**, 3653 (2016).
[25] N. Kaiser, *Astrophys. J.* **439**, L1 (1995).
[26] P. Schneider, *Mon. Not. R. Astron. Soc.* **283**, 837 (1996).
[27] F. Bernardeau, *Astrophys. J.* **433**, 1 (1994).
[28] F. Bernardeau, *Astron. Astrophys.* **301**, 309 (1995).
[29] R. Juszkiewicz, F. R. Bouchet, and S. Colombi, *Astrophys. J. Lett.* **412**, L9 (1993).
[30] F. Bernardeau, L. Van Waerbeke, and Y. Mellier, *Astron. Astrophys.* **322**, 1 (1997).
[31] R. G. Crittenden, P. Natarajan, U.-L. Pen, and T. Theuns, *Astrophys. J.* **568**, 20 (2002).
[32] S. R. S. Varadhan, *Probability Theory* (AMS, Providence, 2001).
[33] Wolfram Research, Inc., *Mathematica 9.0* (Wolfram Research, Inc., Champaign, Illinois, 2012).