Theoretical modelling of non-linear effects on the statistics of weak gravitational lensing fields

Alexandre Barthelemy

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS
DEPUIS 1257

THÈSE DE DOCTORAT
DE SORBONNE UNIVERSITÉ

Spécialité : Astronomie & Astrophysique

École doctorale n°127: Astronomie et Astrophysique d'Île de France

réalisée à l'Institut d'Astrophysique de Paris

sous la direction de Sandrine Codis, Francis Bernardeau & Raphaël Gavazzi

présentée par

Alexandre BARTHELEMY

pour obtenir le grade de :

DOCTEUR DE SORBONNE UNIVERSITÉ

Sujet de la thèse :

"Modélisation théorique d'effets non-linéaires sur la statistique des champs faibles d'astigmatisme cosmique"

soutenue le 14 décembre 2021

devant le jury composé de :

Mme	Camille BONVIN	Rapportrice
M.	István SZAPUDI	Rapporteur
M.	Jonathan BLAZEK	Examinateur
M.	Nicholas KAISER	Examinateur
M.	Éric JULLO	Examinateur
M.	Michael JOYCE	Président du jury
Mme	Sandrine CODIS	Encadrante de thèse
M.	Francis BERNARDEAU	Encadrant de thèse
M.	Raphaël GAVAZZI	Encadrant de thèse

À Alice, je passe devant pour
ne pas souffrir de la comparaison.

Table of contents

Acknowledgments

It seems to me that it is very hard to write a meaningful acknowledgements section. I do not want to sound too mellow, after all I hope everyone involved with me or my work knows how much they mean to me, nor do I want to make anybody mad for not mentioning them, or worse, mentioning something they do not like! So with a mix of sugar and salt here is my best try!

My first thanks must go to my supervisor Sandrine Codis. Sandrine, I enjoyed every minute discussing and working with you, you are the most brilliant researcher I know and truly an inspiration and a model. I have learnt so much, you have made my stay at IAP very smooth and most of all, you have endured my silliness and infuriating sense of humour, sometimes even partaking in it. I know that we often joke about having to quarrel because it looked like every successful researcher in our surroundings seemed to have gotten into a fight with their supervisor, but I do not think I will, I would miss Noélie's tattoos too much! Over the course of this thesis, you have been a mentor, a colleague and now a friend, I hope we will continue to work together for a long time! (Also the food at AIM is much better than at IAP so I want to come back).

To Francis Bernardeau, I must also address some very warm thanks. Francis, you too have endured my sense of humour, but though it appeared to sometimes tire you, you kept coming back for more! I know that deep down you liked teasing me as much as I you. In terms of what you gave me, if Sandrine is truly a rising star of our field, then you have clearly reached the top. You always have insanely good remarks and ideas, your mastery of so many concepts and ideas is impressive and it has been an indescribable pleasure to work with you too! I know you have given up some of your awake time to become the Euclid deputy-lead and thus will have less time for science (it seems very political, not unlike another famous Francis), but you know that I will come back to you for advise and collaborations!

Last but not least among my supervisors, I want to sincerely thank Raphaël Gavazzi. Though you clearly and voluntarily abandoned me to go to Cambridge and then Marseille (do not get mad, I know that is not true), you have been there all along the thesis for moral support and (relevant) future career advise. I unfortunately did not take enough time to work on more science projects with you, which I kind of regret, but there will be more opportunities in the near future! You nevertheless helped me every time I reached out and your scientific culture is extremely impressive. I hope you enjoyed our discussions too, especially when I was trying to explain the hard problems that were on my head at the times, often forgetting to remind you the context :)

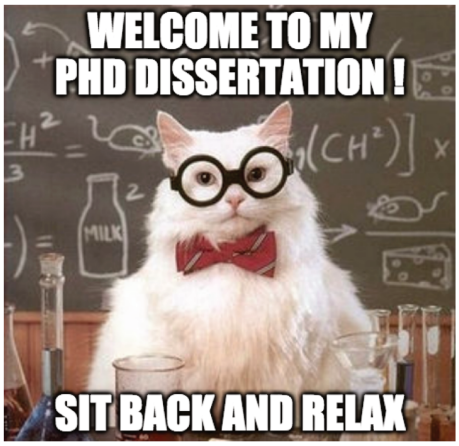
I also want to have a word for my collaborators from the "PDF team", Aoife Boyle, Cora Uhlemann, Oliver Friedrich and Alex Gough. It is so nice to meet and discuss every week with all of you, you set the bar very high for me to perform as well as you all do and you always have a good remark when you are presented with a recent challenge! It is also very enlightening to hear all of your different ideas to push our field forward and I hope we will continue our collaboration for some time! Thankfully I am even joining Oliver for my post-doctoral position in Munich, I am sure we will do great things together there! Aoife, it is always so nice to share a cab with you when we go to CEA, I so enjoy working with you,

talking about politics, music, anything, I will miss you in Munich and hopefully will come back regularly to visit! I also want to thank you for checking my English throughout the manuscript, I swear mentioning this is not a way to dilute my responsibility if we missed any typo ;)

To all of my fellow and ex PhD candidates at IAP, I address the warmest of thanks! For all the table tennis matches and tournaments, the nights watching movies together, the drinks, the restaurants and the coffee breaks, it truly feels like a weird family of science geeks! Of course special thanks go to some of you that will remain some of the best friends I made during my time here: Pierre, Arno, Lukas, Aline, and the others, I love you all!

To Alice, my now fiancée and the love of my life, you have supported me so much during those three years, just as you have supported me before and (hopefully) how you will support me after. This work is dedicated to you, not that it is good enough – that's why I also bought a ring just in case – but it would never have been possible without your company.

Of course, and as the tradition demands, my last words are for my family, my parents which I hope are proud (and now used to the fact that I will try my best not to be an engineer again). Thank you for having provided me with the necessary education so that one day I could be able to write this! You might actually be entitled to a refund, I'll contact my old high-school. Thanks also to my brother, you are so smart and capable, thank god I am older so that I can still have some years of peace before you steal my thunder.



Introduction

General context

We are currently at the frontier between two eras in the field of cosmology. We are at the end of one because the last couple of decades have seen the emergence of a remarkably accurate concordance model in agreement with numerous observations which thus explains much of the physics of our Universe, and whose parameters are also well constrained and roughly consistent in-between different experiments probing different epochs and scales of our Universe. On the other hand, we are also at the beginning of another era because no new fundamental theory nor approach has yet emerged from the pool of all the discussed theories to successfully explain the discrepancies and fundamental issues that we currently face. Waiting for (and working towards) this new paradigm, our actual one relies on the so-called ΛCDM model plus an inflationary phase which give us an overall history of our Universe that I now briefly describe.

In this model, the Universe initially emerges as an initial hot plasma out of a Big-Bang singularity 13.8 billion years ago, and expands which allows the plasma to cool down. Almost immediately after ($t \lesssim 10^{-35}$ seconds) this un-explained singularity (it is after all a consequence of rewinding the model back in time rather than a forward prediction), a period of cosmic inflation renders the Universe flat and homogeneous (Guth, 1981), followed by the separation of the electroweak interaction into electromagnetism and the weak interaction (electroweak phase transition roughly $\lesssim 10^{-11}$ seconds after the Big-Bang). As the Universe cools down, the initial plasma of elementary particles (quarks and gluons) transitions into hadrons, this is called the quark-hadron transition and happens, still in this model, $\lesssim 10^{-6}$ seconds after the Big-Bang. The Universe is then dominated by photons and enters what is known as the radiation-dominated era. Note that these first steps of the Universe evolution are very speculative since current physical theories lack a complete and accurate description of what happens at very high energies ($>$ GeV). The following steps of the Universe evolution have their own but very different difficulties. Indeed they may be very complex, but they nevertheless take place in regimes for which we understand the fundamental laws of physics. A few minutes (between 3 and 20) after the Big-Bang, the primordial nucleosynthesis allows the formation of the first light atomic nuclei like Hydrogen, Deuterium, Helium, Lithium and Beryllium. Since the photon density decreases faster than the matter density (a^{-4} versus a^{-3} if a is the scale factor accounting for the expansion of the Universe), their densities become equal after 70,000 years and we enter the so-called matter-dominated era where dark matter

perturbations start to grow. After a while, free electrons are captured by the light atomic nuclei and neutral hydrogen is created which is a phenomena known as the recombination. At the end of this process, photons decouple from matter and can then travel freely through the Universe which thus becomes transparent. This first light was emitted 380,000 years after the Big-Bang and is now observed, because of the expansion of space-time, as a Cosmic Microwave Background (CMB) which has been extensively studied since its discovery in 1965. This observation is at the forefront of hints toward a hot Big-Bang scenario. The CMB is observed to follow an almost perfect black body emission and is currently measured at a temperature of 2.725 K in all directions in the sky up to tiny fluctuations of order 10^{-5} which are believed to come from initial quantum fluctuations and to be the seeds for the formation of the large-scale structure of our Universe. This makes the Universe at the time of recombination very homogeneous and isotropic. However, though the Universe was transparent, the clouds of hydrogen only collapsed very slowly to form stars and galaxies which makes the only photons present at this time those emitted at the end of recombination. This is known as the dark ages. After the dark ages, from 300 million to one billion years after the Big-Bang, the first stars and galaxies form and a reionisation epoch begins since the intense radiation they emit reionises much of the Universe splitting the neutral hydrogen atoms back into a plasma of free electrons and protons for the first time since recombination. The ionised hydrogen gradually reverts to neutral atoms one billion years after the Big-Bang when the era of the earlier stars and quasars (and their intense radiation) comes to an end. After nine billion years, the Universe starts to be dark energy dominated and therefore undergoes a late-time period of accelerated expansion. The large-scale structure continues to form on larger and larger scales and is another paramount piece of evidence of the cosmological model.

Formally, the ΛCDM model is based on an application of General Relativity at large scales in our Universe then described by an homogeneous, isotropic, flat and expanding space-time whose metric is formally given by the Friedmann-Lemaître-Robertson-Walker metric. For smaller scales, the metric requires the addition of perturbations that describe more precisely the matter field and gravitational waves. The model also requires the introduction of dark energy through the constant Λ entering the Einstein equations in order to explain the acceleration of cosmic expansion in the last four billions years. Another component, collisionless cold – non-relativistic – dark matter (CDM) is needed to explain the growth of structure, and in particular the population of galaxies that we observe today. Note that this dark matter component does not interact electromagnetically and therefore can not be directly detected. Only indirect detections exist with the first one reported by Zwicky in 1933 studying the dynamics of the Coma cluster and arguing that those could only be explained if it was 300 times more massive than what observations suggested. The study of the rotation curves of galaxies by Rubin in 1980 were also paramount in bringing to light the dark matter halos around them. Dark matter fluctuations can moreover start to grow as soon as the Universe becomes matter-dominated which then allows baryonic matter, once decoupled from photons, to fall into the gravitational potential wells previously produced by dark matter thus enabling the formation of structures such as galaxies. Eventually, gravitational lensing and the CMB put tight constraints on the amount of dark matter in the Universe. The nature of dark matter is still unknown and though numerous candidates exist (weakly interactive massive particles like neutrinos or neutralinos, axions, Kaluza-Klein

particles), none has yet been detected. As for dark energy, whose necessity was discovered only two decades ago by Riess et al. (1998); Perlmutter et al. (1999), its nature is still completely unknown: real constant in Einstein field equations, vacuum energy, dynamical scalar field, hint for a modification of gravity, there are numerous possibilities which question our current understanding of the physics of our Universe and are at the heart of current research in cosmology.

In any case and within the ΛCDM paradigm, the energy content of the Universe today is constrained to be roughly 69% in dark energy, 26% in dark matter and 5% in baryons (Planck Collaboration et al., 2020b). Radiation only accounts for 10^{-5} of this energy budget and the minimal number of parameters that describe the ΛCDM cosmological model is six, three for the energy content: densities of baryons, dark matter and dark energy, two for the initial power spectrum of fluctuations: the initial scalar spectral index n_s and the curvature fluctuation amplitude A_s, and the reionisation optical depth τ. Again remember that this model assumes that the spatial curvature is zero. To these base parameters, we often add the value of Hubble parameter today $H_0 \equiv \dot{a}/a \equiv 100h \approx 67$ km s^{-1} Mpc^{-1} (1 Mpc $\approx 3.086 \cdot 10^{22}$ m), the amplitude of linear matter fluctuations in a sphere of radius 8 Mpc/h, $\sigma_8 \approx 0.81$, and eventually the sum of neutrino Masses $M_\nu \approx 0.06$ eV. The sum of neutrino masses is not technically part of the ΛCDM paradigm but even assuming ΛCDM, neutrinos have a distinctive signature in data coming from the fact that they start as relativistic particles and slow/cool down to behave as regular non-relativistic matter.

The ΛCDM model is thus relatively simple but is now well-constrained and has had many impressive successes. As previously mentioned we indeed predict and observe with incredible accuracy the black body radiation of the cosmic microwave background and the spectrum of its anisotropies (CMB Planck Collaboration et al., 2020a), predict the creation and abundance of primordial elements (Fields et al., 2020), can predict and observe the distribution of galaxies and the correlation of their shape due to gravitational lensing (The Dark Energy Survey Collaboration, 2005, this is of course a general picture which will be nuanced), reproduce the light curves of supernovaes which act as standard candles to measure radial distances in the Universe (Betoule et al., 2014), and reproduce the baryon acoustic oscillations which act as a ruler for transverse distances in the Universe (Alam et al., 2017). All of this thus looks like the end of an era for cosmology with the strong establishment of the concordant standard model of cosmology, and the beginning of new one, now focused on finding hints of new physics in the physics of inflation, more theoretically motivated justifications for dark energy and dark matter, and even explain some tensions that have appeared as the constraining of cosmological parameters became more and more precise. Indeed, recent results point towards a disagreement in the value of the Hubble parameter H_0 inferred from local, late-time measurements (Riess et al., 2018; Yang et al., 2020) and from CMB studies (Planck Collaboration et al., 2020b). Still from CMB studies, there seems to be a departure of measurements from the ΛCDM prediction in the CMB spectrum at very large angular scales (Planck Collaboration et al., 2020a), and there is also a relative tension on σ_8 and Ω_m again from the CMB and late-time probes like gravitational lensing or cluster counts (Battye et al., 2015). There are also some difficulties on smaller scales where baryonic physics, which is however complicated to model, becomes important, for example the missing satellite problem where N-body simulations predict much more dark matter substructures than the observed number of galactic satellites, or the cusp-core problem where dark matter

simulations predict a central cuspy density profile for dark matter halos while observations find almost flat density profiles in the inner regions of galaxies. Note that these last issues are with more difficulty attributed to pure inaccuracies in the cosmological model as they could also be resolved with astrophysical arguments.

Statistics of large-scale structure

Though we could think true that the Universe is indeed homogeneous and isotropic on large scales, we know at least since 1986 and the CfA Redshift Survey (de Lapparent et al., 1986) that there exists superstructures (clusters of galaxies, superclusters, with filaments and walls connecting them, and voids in between) on scales between a few tens and hundreds of Mpc, and I have already mentioned that those structures are believed to originate from initial quantum fluctuations in the matter density field. Those initial density perturbations then grow under the laws of gravity and hierarchically form galaxies, clusters and super-clusters of galaxies. As such, be it for the study of the early Universe, or its subsequent late-time evolution, the search for solutions to some issues of the ΛCDM paradigm can be performed in the detailed study of the large-scale structure of our Universe through the use of galaxy surveys that catalogue the distribution of baryonic matter, and even dark matter thanks to gravitational lensing. Indeed, in the forthcoming years, several huge experiments, so-called stage IV galaxy surveys, will be launched and will start to acquire an unprecedented amount of data of no less unprecedented quality. I can notably name the Euclid (Laureijs et al., 2011) and the LSST (Ivezić et al., 2019) experiments, the former having financed my work over the course of the last three years.

Now, to obtain more and more precise constraints on the values of the cosmological parameters and the dark energy equation of state in order to elucidate its true nature from this huge amount of data, we first need to correctly derive the complicated non-linear evolution of structure formation within given models so that we can link what we observe to precise estimation of the parameters. This is in general done by several means. In the current paradigm, the growth of cosmic structures is driven by the gravitational instability meaning that the densest regions attract matter from the rarest. The starting point of this process is then the quasi-homogeneous and quasi-Gaussian primordial density field as found at the redshift of the CMB, and the evolution of these primordial fluctuations could be done through Einstein's equations for an homogeneous universe perturbed by small density fluctuations δ. Einstein's equations are highly non-linear but as far as sub-horizon modes are concerned, the problem can be reduced to its Newtonian approximation leading to the Vlasov-Poisson system and perturbation theory can be applied for density contrasts much smaller than unity. Several perturbation strategies exist, in Eulerian or Lagrangian space (Bernardeau et al., 2002) which have different (rather poor) convergence rates, or with renormalisation procedures to ensure better convergence (Effective field theory Carrasco et al. (2012), RegPT Taruya et al. (2012) etc). I will not present those aspects in detail and though it is not mandatory to understand what is presented in this thesis, I will moreover assume some basic knowledge of Eulerian perturbation theory, for example the notion of perturbation theory kernels or the basic computation of the third cumulant of the density field. Bernardeau et al. (2002) is a very complete reference to understand those aspects. Another mean to access

the non-linear evolution of the density field is via numerical simulations with or without baryonic processes, these are presented in a bit more detail in section 1.8.

Finally, the knowledge of the non-linear density field must be translated into specific observables in the survey, for example in Fourier space of real space. The most common being the so-called poly-spectra (Peebles, 1980) and more precisely the 2-point correlation function and sometimes the 3-point function. This comes from the fact that the 2-point correlation function contains complete statistical information for Gaussian fields, as is the case at least in the linear regime of structure formation starting from the Gaussian primordial fluctuations in the CMB. Let me take the time to briefly introduce those. Given a classical random field such as the matter density contrast $\delta(\boldsymbol{x})$ and under the ergodicity hypothesis which replaces ensemble averages over different realisations of the Universe by spatial averages, the expectation value of any function f of the density field is given by

$$f_0 = \langle f(\mathbf{x}) \rangle = \frac{1}{V} \int f(\mathbf{x}) \mathrm{d}^3\mathbf{x}. \tag{1}$$

The statistics of the random field can then be described by its successive N-point correlation functions. If the field is for example homogeneous and isotropic, the two-point correlation function $\xi(r) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle$ only depends on the magnitude r of the separation $\boldsymbol{r}$. In Fourier space this reads

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle = \delta_D(\mathbf{k} + \mathbf{k}') P(k), \tag{2}$$

where $P(k)$ is the power spectrum and the Fourier transform of $\xi(r)$. In higher dimensions, the connected successive correlations functions between n planar wave modes define the successive poly-spectra P_n

$$\langle \delta(\mathbf{k}_1) \dots \delta(\mathbf{k}_n) \rangle_c = \delta_D(\mathbf{k}_1 + \dots + \mathbf{k}_n) P_n(\mathbf{k}_1, \dots, \mathbf{k}_n). \tag{3}$$

In this thesis, I am rather focused on a specific observable, the Probability Distribution Function (PDF, see appendix A.1.2) of the lensing fields. This is not the most common observable but I will try to convey how important it can be, both taking into account the non-Gaussianities that the density field develops through the gravitational instability and thus containing valuable cosmological information.

Motivation for my work

The general context in which my work takes place is thus the development of accurate models for statistics of the observed fields in the large-scale structure of our Universe, this in preparation of up-coming large galaxy surveys. These serve to constrain more and more the standard cosmological model and its direct extensions to, all other things being equal, elucidate the mysteries surrounding dark energy and dark matter. In this context of need for very accurate predictions, from-first-principles theoretical models driven by perturbation theory and its cousins are not very popular and a large part of the community rather works with more phenomenological models and numerical simulations that have the advantage of *a priori* probing more accurately the non-linear regime of structure formation. However, being able to probe the non-linear purely gravitational evolution of structures hides the fact

Ω_m	Ω_Λ	$\Omega_{\rm cdm}$	Ω_b	h	σ_8	n_s
0.279	0.721	0.233	0.046	0.7	0.82	0.97

Table 1: Cosmological parameters used throughout the majority of this thesis, except in section 3.8 related to the Fisher forecast of the convergence PDF. They are consistent with the WMAP 9 year result and are used in the lensing simulation Takahashi et al. (2017) from which I extract the majority of the simulated results presented here.

that baryonic processes, not yet well modelled even in hydrodynamic simulations, have a strong influence on the statistics of the fields. Moreover, accurate simulations for various cosmological models are quite expensive and are also prone to convergence issues which necessitates a minima code comparisons for every observable of interest. On the other hand, there exist theoretical tricks to track the non-linear regime with some accuracy, for example looking at spherically symmetric observables as I do in this thesis, and the advancements of numerical simulations allow us to probe the validity regime of these theoretical approaches. As a consequence, the cosmological dependence of observables modelled from first theoretical principles is much more apparent and the evaluation of their information content can be performed without fear of sometimes uncontrolled numerical inaccuracies. All of these points were found to be true for the case of the matter density field as modelled with large deviation theory (Bernardeau & Reimberg, 2016) and I present here the extension of this formalism to observables coming from weak gravitational lensing experiments. In all the computations presented here (except section 3.8) I will assume a ΛCDM universe with cosmological parameters given in table 1.

Papers from which this thesis is built, and personal contributions

Most of the results presented in this thesis are taken from a set of five papers in which I was involved. I led four of them for which I am first author and performed most of the work on my own, while the other is a more collaborative work from the small "PDF with large deviations" team that came to life over the course of my PhD and that now meets regularly to discuss and build new projects. I do not describe in detail the content of each paper – this is done in the relevant chapters – but rather focus on the role and contributions that I had in each of them. Most of the figures and tables in this thesis are published in the papers relevant to the sections in which they are. This is always indicated in the caption.

Barthelemy et al. (2020a): This paper presents the first results I obtained modelling the convergence CGF/PDF with large deviation theory and combining it with the BNT transform. I performed every computation, implementation and plots present in this paper as well as wrote it in its entirety. The role of the other authors, mainly my thesis supervisors, was to proof-read what I wrote and discuss all the obtained results with me. My supervisors also gave me the original idea. The last author, Cora Uhlemann, participated in the discussions and had the idea, as well as implemented it in parallel, of the analytical shortcut presented in appendix F on the paper. Overall this paper is described in details in chapter 3.

Barthelemy et al. (2020b): This paper presents a detailed derivation of the post-Born terms and the induced correction of the convergence skewness as presented in chapter 1. It also introduces the method to organically include these corrections within the large deviation formalism for the convergence presented in the previous papers. This method is described in chapter 3. Once again, and this true of all my first authored papers presented in this thesis, I did all the computations, implemented everything myself, produced all the plots and wrote the paper in its entirety. Of course some of the derivations I present could already be found in the literature, for example the post-Born terms on the convergence, but I am just pointing out what my role was in each paper. I had the original ideas for the different implementations of the post-Born corrections to the PDF, and the very idea that those corrections could matter at high redshift came during a discussion at the time of the first paper when I performed some comparison of the formalism to the simulated convergence PDF at the redshift of the CMB.

Barthelemy et al. (2021b): This paper presents the extension of the large deviation formalism for the convergence to the more observable aperture mass. It is described in detail in the first part of chapter 4. It also shows the presence of some inconsistencies in the measurements of certain subtle non-linear effects in high-order statistics in the numerical simulation I am using. This calls to caution and in some sense supports the need for accurate theoretical predictions of these observables. I wrote and performed all derivations, measurements and plots in this paper. The other authors, my thesis supervisors, helped a lot during long discussions to test the issues I found in the numerical simulation.

Boyle et al. (2021): This paper presents the first cosmological forecast of the information content of the convergence PDF as modelled by our theoretical formalism. The covariance matrix is taken from fast log-normal simulations and is found to be very accurate. Our Fisher forecasts establish that the constraining power of the convergence PDF compares favourably to the two-point correlation function for a Euclid-like survey area at a single source redshift. When combined with a CMB prior from Planck, the PDF constrains both the neutrino mass M_ν and the dark energy equation of state more strongly than the two-point correlation function. This paper was led and written by Aoife Boyle (first author). I nevertheless wrote the appendices on the large deviation principle and the effective mapping method, as well as part of section 2 on the theoretical model for the PDF which is the one presented in **Barthelemy** et al. (2020a). I performed the plots of figures 1 to 5 and provided the codes and/or data so that they could be reproduced by the main author to better fit the aesthetics theme of the plots in the paper. Generally speaking, I provided my own implementation of the theoretical model for the PDF and helped the main author to reproduce it. I was also an active member in the discussions we had every week to discuss the results and advancement of the paper.

Barthelemy et al. (2021a): This paper exploits the particular correlation structure of BNT observables along the line of sight in the context of a tomographic analysis. Based on this peculiar structure, I present several strategies to compute the joint PDF of these BNT observables in a very efficient way. This is presented in detail in the second half of chapter 4. The idea to exploit the decimation strategy to efficiently compute the joint PDF comes from Francis Bernardeau (second author) with whom I in parallel performed the

derivations presented in this paper. I also wrote it and produced the different figures. The appendices were written by Cora Uhlemann (last author) who had the idea of expressing the problem in terms of mathematical copulas to provide a useful approximate to the exact computations I performed. The remaining author, Sandrine Codis, helped proof-read the article and generally speaking discussed with me every result I produced.

1

Basics of weak gravitational lensing

Contents

1.1. Introduction

There exist a number of powerful probes of the large-scale structure of our Universe: galaxy clustering, cluster number counts, Sunyaev-Zel'dovich effect etc.* However these aforementioned methods all rely on measures of the distribution of galaxies or more generally baryonic matter as opposed to the distribution of the total mass, for which it is much easier with our current theoretical tools to get accurate predictions. Though it is possible to try and establish an effective connection (only at large scales) between galaxies and matter (galaxy bias, Desjacques et al. (2018)), current developments can only be trusted on large scales and at the price of introducing additional free parameters that come from this galaxy bias expansion.

On the other hand, light rays coming from distant sources and that propagate through the inhomogeneous distribution of both baryonic and dark matter are distorted on their way to their observer. This effect, known as gravitational lensing, is very promising because the light path responds to mass, or more precisely, to the spacetime perturbations on top of the homogeneous cosmological background. This technically means that if we were to measure these distortions, then we might be able to infer some information on the distribution of the mass in the universe, or at least between us and the sources we observe. This property thus makes gravitational lensing highly complementary to the other probes of large scale structure.

This idea of gravitational lensing can be traced back as far as the idea of general relativity itself, at least for a rigorous treatment, with early notebooks of Einstein containing calculations of the magnification of images and even of the possibility of multiple images of single sources. Let us not forget, moreover, that it was the 1919 Eddington experiment – a measure of the deflections of starlight by gravitational lensing during a solar eclipse – that led to a more general acceptance of general relativity.

*Though not described in this thesis, these probes are very well introduced in Dodelson, S. and Schmidt, F. (2020) while a recent joint analysis of galaxy clustering and lensing can be found in DES Collaboration et al. (2021).

Nowadays, gravitational lensing is used in multiple contexts, some of which are useful for cosmology. There is for example time delay, where two light rays emitted at the same time by the same source are detected coming from different directions and at different times due to lensing, with the time difference directly depending on the Hubble constant H_0 (Wong et al., 2020). Microlensing, when a lens moves into the line connecting a source to us and thus produces a characteristic variability in the source as a consequence of being magnified, has served for example to constrain the contribution of massive halo objects (MACHOs) to dark matter (Tisserand et al., 2007). Thanks to magnification, lensing also enables the detection and study of galaxies behind massive lenses that would otherwise be too faint to be detected (Coe et al., 2019).

For cosmology, one important aspect of gravitational lensing is known as *weak lensing* (Kilbinger, 2015), where the shapes of distant galaxies are very slightly distorted* by the foreground matter overdensities. I depict the effect in figure 1.1, with the notation being described in the rest of this text. In this chapter, I will only be interested in weak lensing by the large scale distribution of matter in the universe rather than by distinct identifiable lenses such as galaxy clusters. My goal is to present an almost complete derivation of the formalism starting from first principles while trying to present it through an angle that differs a bit from that found in most textbooks on the matter. Hence I will start by deriving the well-known optical Sachs equation, present a method to obtain its solutions in a perturbative framework, and obtain the traditional lensing formalism that considers only terms where lenses are statistically independent. I will make the link between this formalism and the ellipticities we observe, comment on systematics not directly coming from the formalism but that still affect our observations, and introduce the necessity as well as limits of numerical simulations and their complementarity with theoretical approaches. As an exercise I will then present a complete derivation of some next-order corrections to the usual approximations to account for some couplings between lenses, and I will end this chapter with the presentation of the Bernardeau-Nishimichi-Taruya (BNT) transform, a trick that enables accurate theoretical predictions for weak-lensing observables. Note that this chapter, and for that matter everywhere until stated otherwise, focuses on establishing correct predictions for the lensing fields and their statistics in a given cosmological model. In the end we aim at doing the opposite, that is, we infer a cosmological model based on the measured statistics of those fields. I will introduce this in more detail in section 3.8 where, from the then-established model of the convergence PDF, we will infer the constraint on cosmological parameters one could expect. Many derivations in this chapter, especially of post-Born corrections, were published as part of **Barthelemy** et al. (2020b), which includes their effect on the convergence PDF.

1.2. Sachs equation

The weak lensing formalism aims at computing the small deformations undergone by images of distant objects, *e.g.* galaxies, and thus the so-called 2x2 deformation matrix $\mathcal{D}_{ab}$, which is the Jacobian of the transformation from the actual physical separation between two points – the thickness of the light beam – and its perceived angular thickness by the observer. Said differently, this matrix indeed links the infinitesimal separation between two nearby null

i.e. the effect is only detectable through shape correlation of close galaxies.

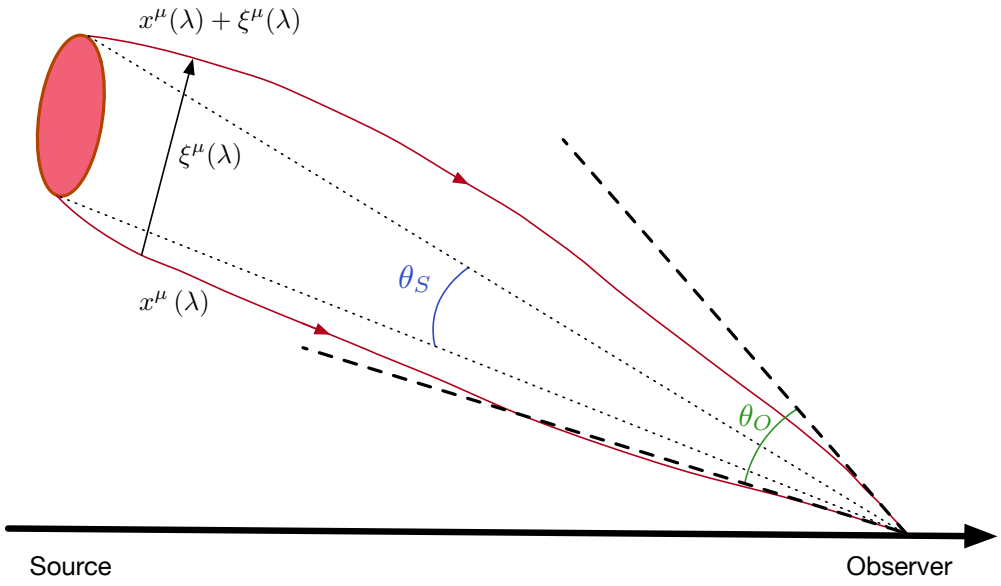


Figure 1.1: Schematic view of possible trajectories of photons coming from a source to an observer. Quantities of interest are labelled as in the main text. This figure was published in **Barthelemy** et al. (2020b).

geodesics ξ^a - we are talking about very thin light beams being deformed - and the vectorial angle seen by the observer between the two geodesics θ_O^a. If $a(z)$ denotes the cosmological scale factor, we can express the (now comoving) deformation as:

$$\xi^a = a(z)\mathcal{D}_{ab}\theta_O^b. \tag{1.1}$$

Note that the vectors ξ^a and θ_O^a are two-dimensional since they are expressed in the basis of the *screen* that the observer looks at.

What is now needed is an evolution equation for $\mathcal{D}_{ab}$, which is given by the so-called optical Sachs equation (Sachs, 1961). It formally reads

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}a(z)\mathcal{D}_{ab} = a(z)\mathcal{R}_a{}^c\mathcal{D}_{cb}, \tag{1.2}$$

where $\mathcal{D}_{ab}(0) = 0$ and $\mathrm{d}\mathcal{D}_{ab}/\mathrm{d}\lambda(0) = \delta_{ab}$ with δ_{ab} the identity matrix. These conditions express that the geodesics are focused at the observer and that spacetime near the observer is Euclidean, *i.e.* no deformation occurs right before the light reaches the observer. Here λ parametrises the geodesics in such a way that we assume the same value at the observer, $\lambda_O = 0$, and can be thought of as a clock, and the optical tidal tensor $\mathcal{R}_{ab}$ is expressed from the Riemann curvature contracted with null vectors k^ν and screen basis-vectors n_a^μ and n_b^σ,

$$\mathcal{R}_{ab} \equiv R_{\mu\nu\rho\sigma}k^\nu k^\rho n_a^\mu n_b^\sigma. \tag{1.3}$$

Because of the great importance of this equation, which I personally think of as the weak-lensing "master equation", and in an attempt to make it clearer and more accessible, I will now dedicate the rest of this section to its formal derivation before describing its resolution in the cosmological context in the following sections. Some elements that I present here were

found in Bernardeau et al. (1997) and Bernardeau et al. (2010) and I add some personal details for the sake of clarity. Finally, I would like to emphasise the fact that the derivation I here propose makes use of the usual thin light-beam approximation which thus prevents the description of lensing events such as multiple images of the same source that do not occur in the context of cosmological weak-lensing. However, it is remarkable that this approximation can be lifted as was done in Fleury et al. (2017, 2019a,b), and although this does not change the validity of our more usual weak-lensing derivations, it does make a more formal link between the so-called weak and strong (which for example allows for multiple images) lensing formalisms.

Here the derivation is proposed using Einstein's summation convention, the $(-1,1,1,1)$ signature and natural units where $c = H_0 = 1$. First, let me consider two nearby geodesics $x^\mu(\lambda)$ and $x^\mu(\lambda) + \xi^\mu(\lambda)$ that lie in the past light cone of an observer O and which are connected by a vector field ξ^μ. As mentioned before, λ parametrises the geodesics in such a way that we assume the same value at the observer, $\lambda_O = 0$. This allows us to define the wave vector of the photons $k^\alpha = \mathrm{d}x^\alpha/\mathrm{d}\lambda$, which obeys the geodesic equation

$$\frac{Dk^\mu}{D\lambda} \equiv k^\alpha \nabla_\alpha k^\mu = 0, \tag{1.4}$$

where the covariant derivative along the geodesic $D/D\lambda$ is expressed below in terms of partial derivatives and Christoffel symbols once a coordinate system is chosen. We consider rays with infinitesimal separation and thus the separation vector lies on the null surface $\xi^\mu k_\mu = 0$ everywhere along the geodesic. In the following, I will derive the evolution equation for the separation vector field ξ^μ. First let us note that given an arbitrary coordinate system,

$$\begin{aligned}
\frac{D\xi^\mu}{D\lambda} &= k^\nu \nabla_\nu \xi^\mu, \\
&= \frac{\mathrm{d}}{\mathrm{d}\lambda}\xi^\mu + \Gamma^\mu_{\alpha\beta} k^\alpha \xi^\beta, \\
&= k^\mu(x+\xi) - k^\mu(x) + \Gamma^\mu_{\alpha\beta} k^\alpha \xi^\beta, \\
\frac{D\xi^\mu}{D\lambda} &= \xi^\nu \partial_\nu k^\mu + \Gamma^\mu_{\alpha\beta} k^\alpha \xi^\beta = \xi^\beta \nabla_\beta k^\mu,
\end{aligned} \tag{1.5}$$

where we go from the first line to the second by expressing k^ν as a function of coordinates and recognising a total derivative, from the second to the third using the definitions of ξ and k, and from the third to the last line using the definition of the derivative. Using the obtained property, let us now obtain an equation of motion for the separation vector:

$$\begin{aligned}
\frac{D^2\xi^\mu}{D\lambda^2} &= k_\alpha \nabla^\alpha \left(\frac{D\xi^\mu}{D\lambda} \right) = k_\alpha \nabla^\alpha \left[\xi^\beta \nabla_\beta k^\mu \right], \\
&= k^\alpha \left(\nabla_\alpha \xi^\beta \right) \left(\nabla_\beta k^\mu \right) + k^\alpha \xi^\beta \nabla_{\alpha\beta} k^\mu, \\
&= \xi^\alpha \left(\nabla_\alpha k^\beta \right) \left(\nabla_\beta k^\mu \right) + k^\alpha \xi^\beta \nabla_{\alpha\beta} k^\mu, \\
&= \xi^\alpha \nabla_\alpha \left[k^\beta \nabla_\beta k^\mu \right] - \xi^\alpha k^\beta \nabla_{\alpha\beta} k^\mu + k^\alpha \xi^\beta \nabla_{\alpha\beta} k^\mu, \\
&= \xi^\alpha k^\beta \left[\nabla_{\beta\alpha} - \nabla_{\alpha\beta} \right] k^\mu, \\
\frac{D^2\xi^\mu}{D\lambda^2} &= \xi^\alpha k^\beta R^\mu{}_{\nu\beta\alpha} k^\nu,
\end{aligned} \tag{1.6}$$

where almost everything is obtained by the chain rule, and the last line is obtained using the Ricci identity with $R^{\mu}{}_{\nu\beta\alpha}$ being the Riemann tensor. This equation therefore describes the evolution of a light bundle along its geodesic. For an observer in O with 4-velocity v_O^μ, the angle under which the observer sees the geodesics is given by the components of $\mathrm{d}\xi^\mu/\mathrm{d}\lambda$ orthogonal to both k^μ and v_O^μ and it is consequently convenient to define two space-like vectors n_a^μ, $a = 1, 2$, orthogonal to both k^μ and v_O^μ and orthonormal between them such that $g_{\mu\nu}n_a^\nu n_b^\mu = \delta_{ab}$ where $g_{\mu\nu}$ is the spacetime metric. Moreover the subspace $\{n_1^\mu(\lambda), n_2^\mu(\lambda)\}$ is called the *screen* adapted to the observer. This allows us to define a basis at the observer $\{n_1^\mu, n_2^\mu, k^\mu, v_O^\mu\}$, which is extended to all space-time by parallel transporting n_a^μ and v_O^μ along the geodesic. n_a^μ and v_O^μ thus also satisfy geodesic equations such as equation (1.4). As such, for all λ one can decompose

$$\xi^\mu = \xi^0 k^\mu + \sum_{a=1,2} \xi^a n_a^\mu + \xi^u v_O^\mu, \tag{1.7}$$

and noticing that the components along v_O^μ are given by $\xi^\mu k_\mu$ being equal to 0, what remains is

$$\xi^\mu = \xi^0 k^\mu + \sum_{a=1,2} \xi^a n_a^\mu. \tag{1.8}$$

Plugging this decomposition into equation (1.6) leads to

$$\frac{D^2\xi^\mu}{D\lambda^2} = R^{\mu}{}_{\nu\alpha\beta}\xi^a n_a^\beta k^\alpha k^\nu + R^{\mu}{}_{\nu\alpha\beta}\xi^o k^\beta k^\alpha k^\nu, \tag{1.9}$$

where the second term is null because $R^{\mu}{}_{\nu\alpha\beta}$ is anti-symmetric in $\alpha\beta$ while $k^\alpha k^\beta$ is obviously symmetric. Thus projecting the equation along the spatial basis n_a^μ and remembering that it can enter the derivative because $Dn_a^\mu/D\lambda = 0$, we obtain an evolution equation for the components ξ^a

$$\frac{\mathrm{d}^2\xi^a}{\mathrm{d}\lambda^2} = \mathcal{R}^a{}_b\xi^b, \tag{1.10}$$

where the symmetric tensor $\mathcal{R}_{ab}$, called the optical tidal tensor in what follows, is defined by

$$\mathcal{R}_{ab} \equiv R_{\mu\nu\rho\sigma}k^\nu k^\rho n_a^\mu n_b^\sigma. \tag{1.11}$$

Since equation (1.10) is linear it is clear that there exists a relation between $\xi^a(\lambda)$, $\xi^a(0) = 0$ and $\mathrm{d}\xi^a/\mathrm{d}\lambda(0) = \theta_O^a$ written in the form

$$\xi^a = \hat{\mathcal{D}}_{ab}\theta_O^b. \tag{1.12}$$

Note that θ_O^a is the vectorial angle seen by the observer between the two geodesics. $\hat{\mathcal{D}}_{ab}$ is thus the fundamental object we are looking for in lensing, completely describing what we observe, that is the angle we perceive as opposed to the image on the virtual *screen* moving along the geodesic as defined earlier. As seen earlier, we call it the deformation matrix. ξ being infinitesimal, $\hat{\mathcal{D}}_{ab}$ is the Jacobian of the transformation. Finally, plugging equation (1.12) into equation (1.10) leads to the fundamental Sachs equation describing $\hat{\mathcal{D}}_{ab}$

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\hat{\mathcal{D}}_{ab} = \mathcal{R}_a{}^c\hat{\mathcal{D}}_{cb}. \tag{1.13}$$

In the case of a non perturbed space-time, the deformation matrix in equation (1.12) will correspond to the usual angular distance multiplied by the identity matrix. In the cosmological context, this distance is usually expressed in its comoving version and I thus find convenient to rewrite equation (1.13) as

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}a(z)\mathcal{D}_{ab} = a(z)\mathcal{R}_a{}^c\mathcal{D}_{cb} \tag{1.14}$$

with $\mathcal{D}_{ab}(0) = 0$ and $\mathrm{d}\mathcal{D}_{ab}/\mathrm{d}\lambda(0) = \delta_{ab}$ and where δ_{ab} is the identity matrix. Once again, these conditions express that the geodesics are focused at the observer and that spacetime near the observer is Euclidean.

1.3. FLRW optical tidal tensor

Up to this point no usual weak lensing approximations were made (e.g. weak fields or Born approximations) other than considering an infinitesimally thin light beam. We will however now move to the specific cosmological context and compute the optical tidal tensor $\mathcal{R}_{ab}$ given a linearly perturbed (only scalar perturbations) flat Friedman-Lemaître-Robertson-Walker (FLRW) metric which, given the Newtonian gauge and spherical coordinates, can be expressed as

$$\mathrm{d}s^2 = -(1+2\psi)\mathrm{d}t^2 + a^2(t)(1-2\phi)\ [\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2], \tag{1.15}$$

with $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2$. ψ is the Newtonian gravitational potential and ϕ can be interpreted as some perturbation of the scale factor $a(t)$. Note that the magnitude of the two Bardeen potentials ψ and ϕ is very small for most mass distributions in the universe, excluding only very compact objects whose extent is comparable to their Schwarzschild radius, which justifies the use of linear perturbations excluding couplings between the potentials. According to the definition given in equation (1.3), we need to contract the Riemann curvature tensor at first order in ϕ and ψ with the 4-momentum of the photon k^μ and n_a^μ computed at the level of the background ($\phi = \psi = 0$) so that $\mathcal{R}_{ab}$ is also first order – linear – in ϕ and ψ. The Riemann tensor is straightforwardly computed from the metric so that all that remains is the need to express k^μ and n_a^μ. This can be done noticing that the geodesics of interest are given by $\mathrm{d}\theta = \mathrm{d}\varphi = \mathrm{d}s = 0$, which gives the relation between $\mathrm{d}t$ and $\mathrm{d}r$ and thus between $k^0 \equiv \mathrm{d}t/\mathrm{d}\lambda$ and $k^1 \equiv \mathrm{d}r/\mathrm{d}\lambda$. k^μ is therefore proportional to $(-1, 1/a(t), 0, 0)$ and the proportionality factor comes from the resolution of the geodesic equation. Finally we get

$$k^\mu = \frac{1}{a(t)}(-1, 1/a(t), 0, 0). \tag{1.16}$$

We are relatively free to choose the vectors n_a^μ since only orthogonality conditions must be satisfied. We choose them to be proportional to $(0, 0, 1, 0)$ and $(0, 0, 0, 1)$ where the prefactors are again obtained with the geodesic equation. Thus, the screen basis vectors eventually read

$$\begin{cases} n_1^\mu = \dfrac{1}{a(t)r}(0, 0, 1, 0) \\ n_2^\mu = \dfrac{1}{a(t)r\sin(\theta)}(0, 0, 0, 1) \end{cases}. \tag{1.17}$$

Then plugging equations (1.16)-(1.17) into equation (1.3) we obtain that at first order in ϕ

$$\mathcal{R}_{ab} = \frac{1}{a^2(t)}\left[-(\phi_{,ab} + \psi_{,ab}) - 4\pi G\bar{\rho}\delta^D_{ab}\right], \tag{1.18}$$

where δ^D is the Dirac delta function, the derivatives are taken along the transverse spatial components and where we used the fact that for a flat universe and a pressure-less fluid (Friedmann equations)

$$\frac{\dot{a}^2 - a\ddot{a}}{a^2} = -\dot{H} = \frac{3}{2}H^2 = 4\pi G\bar{\rho}. \tag{1.19}$$

Moreover, in a universe without anisotropic stress, the case considered here and valid at large scales, we also have $\phi = \psi$. Given the relation between the scale factor and redshift together with the comoving Poisson equation

$$\Delta\phi = 4\pi G\bar{\rho}\delta_{\rm mass}, \tag{1.20}$$

where $\delta_{\rm mass} = (\rho - \bar{\rho})/\bar{\rho}$ is the matter density contrast, Δ is the three-dimensional Laplacian operator and derivatives are taken with respect to proper distances, one eventually gets – up to a total derivative in λ with no observational consequences –

$$\mathcal{R}_{ab} = -\frac{3}{2}\Omega_m(1+z)^5\left[\delta^D_{ab} + \phi_{,ab}\right], \tag{1.21}$$

where we redefined ϕ such that $\frac{1}{2}\Delta\phi = \delta_{\rm mass}$. Since this result was computed by linearly perturbing the metric, the potential defined is a reflection of the density contrast computed in the linear regime. Note moreover that for our purposes, that is for the dynamical physical scales and redshifts, Newtonian Eulerian perturbation theory* (Bernardeau et al., 2002) will be sufficient for the description of the higher order density field and thus $\mathcal{R}_{ab}$ can be expanded following

$$\mathcal{R}_{ab} = \mathcal{R}^{(0)}_{ab} + \mathcal{R}^{(1)}_{ab} + \mathcal{R}^{(2)}_{ab} + ... \tag{1.22}$$

$$= -\frac{3}{2}\Omega_m(1+z)^5\left[\delta^D_{ab} + \phi^{(1)}_{,ab} + \phi^{(2)}_{,ab} + ...\right] \tag{1.23}$$

where $\frac{1}{2}\Delta\phi^{(n)} = \delta^{(n)}_{\rm mass}$. Moreover we will suppose that $\mathcal{D}_{ab}$ can also be expanded with respect to $\mathcal{R}_{ab}$. Thus equation (1.2) gives a hierarchy of equations for $\mathcal{D}^{(n)}_{ab}$ that can be solved order by order.

1.4. Deformation matrix up to second order

I aim in this section at solving the Sachs equation (1.2) – that is, to obtain each order of the deformation matrix $\mathcal{D}_{ab}$.

*Newtonian Eulerian perturbation theory is sometimes referred to as standard perturbation theory (SPT) in the litterature.

The zeroth order version of equation (1.2) corresponds to the case of an unperturbed homogeneous universe for which one gets

$$\frac{\mathrm{d}^2[a(\lambda)\mathcal{D}_0(\lambda)]}{\mathrm{d}\lambda^2} = -\frac{3}{2}\Omega_m(1+z)^4\mathcal{D}_0(\lambda), \tag{1.24}$$

whose *only* solution is a well-known result of standard flat FLRW cosmology, that is, the comoving angular distance

$$\mathcal{D}_0(\lambda(z)) = \int_0^z \frac{dz'}{H(z')}. \tag{1.25}$$

Moving to the next order, the first order equation can be written as

$$\frac{\mathrm{d}^2[a(\lambda)\mathcal{D}^{(1)}_{ab}(\boldsymbol{\theta},\lambda)]}{\mathrm{d}\lambda^2} + \frac{3}{2}\Omega_m(1+z)^4\mathcal{D}^{(1)}_{ab}(\boldsymbol{\theta},\lambda) = -\frac{3}{2}\Omega_m(1+z)^4\mathcal{D}_0(\lambda)\phi^{(1)}_{,ab}(\boldsymbol{\theta},\lambda), \tag{1.26}$$

with $\mathcal{D}^{(1)}_{ab}\Big|_{z=0} = 0$ and $\mathrm{d}\mathcal{D}^{(1)}_{ab}/\mathrm{d}\lambda\Big|_{z=0} = 0$. Let me now temporarily make use of the so-called Born approximation by considering that the λ in equation (1.26) corresponds to the unperturbed space-time and we thus have from equation (1.16) that $\mathrm{d}\lambda = (1+z)^{-2}H(z)^{-1}\mathrm{d}z$ since $a(z) = 1/(1+z)$ and $\mathrm{d}r = \mathrm{d}z/H(z)$. This is equivalent to taking every integral along the line of sight following the unperturbed null geodesics. This approximation enables me to avoid the complication of explicitly writing down the parametrisation of the perturbed geodesics but can and will be corrected later on once I will have introduced the convergence field up to second order in section 1.6.

I now express everything in terms of the redshift z rather that λ and notice that the associated homogeneous equation to equation (1.26) is exactly the zeroth order one. As a consequence, its associated Green function $H(z)\mathcal{D}_0(z',z)$ is essentially the angular comoving distance between the redshift of the source and some other redshift, and we finally obtain the first order expression of the deformation matrix by integrating this Green function on the source term

$$\mathcal{D}^{(1)}_{ab}(\boldsymbol{\theta},z) = -\frac{3}{2}\Omega_m\int_0^z \frac{\mathrm{d}z'}{H(z')}\mathcal{D}_0(z',z)\mathcal{D}_0(z')(1+z')\phi^{(1)}_{,ab}(\boldsymbol{\theta},z'). \tag{1.27}$$

The second order equation can be similarly written as

$$\frac{\mathrm{d}^2\left[a(z)\mathcal{D}^{(2)}_{ij}(\boldsymbol{\theta},z)\right]}{\mathrm{d}\lambda^2} - a(z)\mathcal{R}^{(0)}_{ik}(z)\mathcal{D}^{(2)}_{kj}(\boldsymbol{\theta},z) = \\ -\frac{3}{2}\Omega_m(1+z)^4\phi^{(2)}_{ij}(\boldsymbol{\theta},z)\mathcal{D}_0(z) + a(z)\mathcal{R}^{(1)}_{ik}(\boldsymbol{\theta},z)\mathcal{D}^{(1)}_{kj}(\boldsymbol{\xi},z), \tag{1.28}$$

where we once again recognise the zeroth order equation on the left-hand side so that the deformation matrix at second order eventually reads

$$\mathcal{D}^{(2)}_{ab}(\boldsymbol{\theta},z) = -\frac{3}{2}\Omega_m\int_0^z \frac{\mathrm{d}z'}{H(z')}\mathcal{D}_0(z',z)\mathcal{D}_0(z')(1+z') \\ \left[\phi^{(2)}_{,ab}(\boldsymbol{\theta},z') + \frac{\phi^{(1)}_{,ac}(\boldsymbol{\theta},z')\mathcal{D}^{(1)}_{cb}(\boldsymbol{\theta},z')}{\mathcal{D}_0(z')}\right]. \tag{1.29}$$

Note that the structure of the first term in equation (1.29) is very similar to the first order result obtained in equation (1.27), that is, a weighted superposition of second derivatives of the gravitational potential taken at the same order. This contribution exactly corresponds to summing the result of independent single thin lenses along the line of sight using geometrical optics. This is thus exact only at first order. The structure of the second term in equation (1.29) – a double integral – accounts for couplings between the lenses: the effect of one particular lens along the line of sight depends on the effect of all the lenses upstream. In other words, the impact parameter of a photon passing through one lens depends on the source but also on the other lenses the photon passed by. More quantitatively, one can also understand this second term as a generic optical geometric correction whose form appears when one rigorously treats the case of more than one thick lens. Remember that this contribution is at this order only one of the two we should consider since we made use of the Born approximation. The other contribution comes from the fact that lenses themselves are lensed and therefore are not exactly at the angular position they would have been in the absence of lenses. Once again, this could have been directly taken into account if we for example knew how to link the perturbed geodesic parameter to redshift and then had followed the exact same steps as presented here. This will be investigated below in section 1.6. However, let me first, since we have now derived all the necessary material, spend some time on what is considered the usual weak-lensing formalism and that takes place within the Born approximation with independent lenses.

1.5. Weak-lensing within the Born approximation with independent lenses

1.5.1 Standard formalism

The derivation of the deformation matrix leaves apparent that at every order will come a contribution – from the integration along the line of sight – of some projection of second derivatives of the gravitational potential. Since these contributions tend to be dominant*, it is common to define the Jacobian matrix A of the transformation from the source to image plane as

$$A_{ij}(\boldsymbol{\theta}, z) = \frac{\partial \theta_{Si}}{\partial \theta_{Oj}} = \frac{\mathcal{D}_{ij}(\boldsymbol{\theta}, z)}{\mathcal{D}_0(z)} = \delta^D_{ij} - \partial_i \partial_j \psi(\boldsymbol{\theta}, z) \tag{1.30}$$

where ψ is the projected gravitational potential – the *lensing potential* – along the line of sight

$$\psi(\boldsymbol{\theta}, z) = \frac{3}{2}\Omega_m \int_0^z \frac{\mathrm{d}z'}{H(z')} \frac{\mathcal{D}_0(z', z)\mathcal{D}_0(z')}{\mathcal{D}_0(z)} (1 + z')\phi(\boldsymbol{\theta}, z'). \tag{1.31}$$

This definition of the deformation matrix is then established within the Born approximation, and I integrate along the background null geodesics, neglecting all couplings between

*Proper illustration will be given in sections 1.6 and 1.7 below related to the convergence at second order and the computation of the post-Born correction on the skewness.

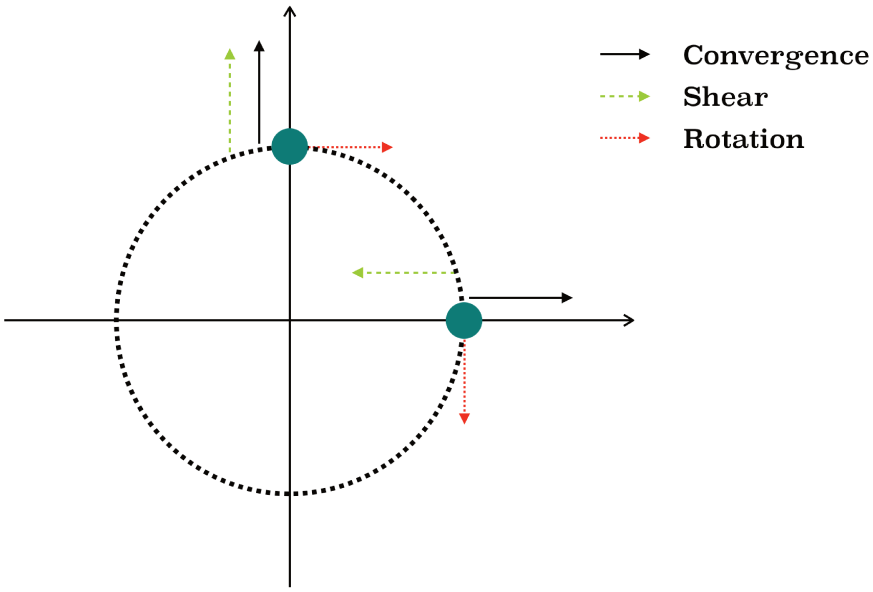


Figure 1.2: Effect of the trace (black), symmetric (green) and anti-symmetric (red) parts of the deformation matrix on points located on the unit sphere.

the lenses as explained when discussing equation (1.29). As with any matrices we can decompose it into a trace, a symmetric and an anti-symmetric part which will all have their own physical interpretations. The trace part will, analogously to the deformation matrix in fluid mechanics, correspond to the isotropic deformation of the light bundle, the symmetric part to its anisotropic stretching or shear and the anti-symmetric part to its rotation. I illustrate the impact of the deformation matrix in figure 1.2 on two unit vectors located on the unit sphere. Note that since I chose to ignore any couplings between the lenses, the deformation matrix is symmetric and thus the rotation part is zero by definition. It is common to parametrise the trace part by the scalar *convergence* κ and the symmetric part by the spin-two* component *shear* $\gamma = (\gamma_1, \gamma_2)$. We thus rewrite A as (Bartelmann & Schneider, 2001)

$$A = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} \tag{1.32}$$

which allows one to more formally define

$$\kappa = \frac{1}{2}\left(\partial_1\partial_1 + \partial_2\partial_2\right)\psi = \frac{1}{2}\Delta\psi; \quad \gamma_1 = \frac{1}{2}\left(\partial_1\partial_1 - \partial_2\partial_2\right)\psi; \quad \gamma_2 = \partial_1\partial_2\psi. \tag{1.33}$$

Remember that in the context of cosmological lensing by large-scale structures, images are very weakly lensed and the amplitudes of κ and $\gamma = \gamma_1 + i\gamma_2$ are of the order of a few percent or less. The inverse Jacobian A^{-1} describes the local mapping of the source light distribution to image coordinates, and each source is mapped uniquely onto one image – we saw that there are no multiple images in this formalism – which thus makes the matrix A invertible. The description of this formalism now brings us to the important question of what is formally observable in an astrophysical survey.

*It is invariant by rotation about π.

1.5.2 Observation of the ellipticities of galaxies

Following the path of Schneider & Seitz (1995) and Seitz & Schneider (1997), it can be shown that for an image with elliptical isophotes, minor-to-major axis ratio b/a, and position angle (or orientation) ϕ, its observed ellipticity $\epsilon = (a-b)/(a+b)e^{2i\phi}$ is given by

$$\epsilon = \frac{\epsilon^s + g}{1 + g^*\epsilon^s}, \tag{1.34}$$

where ϵ^s is the intrinsic ellipticity of the source, $g = \gamma/(1-\kappa)$ is the *reduced shear* and the asterisk denotes a complex conjugation.

Let me sketch rapidly the idea behind the derivation of this result. Let us consider $I_s(\boldsymbol{\theta})$ to be the surface brightness distribution of an extended source, in our case a galaxy. Its matrix of second moments $Q_{ij} \propto \int \mathrm{d}^2\boldsymbol{\theta} I(\boldsymbol{\theta})(\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)$, where $\bar{\boldsymbol{\theta}}$ is the centre of the source, is the appropriate mathematical object with which to describe its shape and spatial extension. The shape, in our case the ellipticity, of the source will hence depend on the eigenvalues of Q_{ij} and it is possible to write ϵ explicitly as a function of elements of Q_{ij}. Now, thanks to the conservation of surface brightness in gravitational lensing – I will come back to this when talking about magnification in section 1.5.6 – the modification of Q_{ij} by the effect of lensing is only a distortion of the image and $Q_{\mathrm{lensed}} = A \cdot Q \cdot A$, which in turn allows us to compute equation (1.34).

Now notice how the ellipticity of the source does not depend on the shear only but on a combination of the convergence and the shear. This could already be seen in the expression of the Jacobian matrix A (1.32). Indeed, since we can factor out $(1-\kappa)$ without affecting the shape of the source, this multiplier only affecting its size, and since our measurement is based on galaxy shapes, then the observable in question is not the shear γ but the reduced shear g. However, in the weak lensing regime where the order of magnitude of $\kappa \approx 10^{-2}$, it is common to approximate the reduced shear $g \approx \gamma$. Now, in the absence of a preferred galaxy orientation we have $\langle \epsilon^s \rangle = 0$* and the ellipcity of galaxies is thus an unbiased estimator of the reduced shear

$$\langle \epsilon \rangle = g \approx \gamma, \tag{1.35}$$

or even the shear in the weak-lensing approximation. This links our mathematical objects to actual observations, and now what remains is to compute the statistics of observable lensing fields and exploit them for tests of our cosmological theories. Starting by writing the formal definitions (1.33) in Fourier space† and using the complex notation for the shear we find that for $\ell \neq 0$

$$\tilde{\gamma}(\boldsymbol{\ell}) = \frac{\ell_1 + i\ell_2}{\ell^2} = e^{2i\beta}\tilde{\kappa}(\boldsymbol{\ell}), \tag{1.36}$$

with β the polar angle of the wave-vector $\ell = (\ell_1, \ell_2)$ written as a complex quantity. This has a number of interesting properties. First of all, it shows that the convergence field can be re-constructed from the observed shear field with reconstruction schemes in the spirit of

*This breaks down in the presence of intrinsic alignments as discussed in section 1.5.4.

†Working in Fourier space rather than spherical harmonics introduces another approximation: the flat sky approximation valid uniquely for small angular scales, that is large ℓ. More generally, this is part of the Limber approximation when computing angular correlators (Lemos et al., 2017).

Kaiser & Squires (1993) that take advantage of this relation, but second, this reconstruction can only always be up to a constant mass sheet κ_0 since the $\ell = 0$ mode is not recoverable from the shear. This last point is known as the mass-sheet degeneracy* (Schneider & Seitz, 1995). Finally, another consequence of equation (1.36) is that the shear and the convergence both have the same power spectrum (Bartelmann & Schneider, 2001).

The convergence power spectrum can then be computed in the following way. Within the assumption that the matter density field on large scales is statistically homogeneous and isotropic, which actually follows from the cosmological principle, the same holds for the convergence. Expressed in Fourier space, these considerations allow us to define the convergence power spectrum P_κ as

$$\langle \tilde{\kappa}(\boldsymbol{\ell})\tilde{\kappa}\left(\boldsymbol{\ell}'\right)\rangle = \delta^D\left(\boldsymbol{\ell}-\boldsymbol{\ell}'\right) P_\kappa(\ell), \tag{1.37}$$

and a similar definition holds for the matter power spectrum P_δ. Now taking the Fourier transform of (1.33) and (1.31), adding them together and taking the average leads to

$$P_\kappa = \frac{9H_0^4\Omega_{\rm m}^2}{4}\int_0^z \frac{\mathrm{d}z'}{H(z')}\frac{\mathcal{D}_0(z',z)^2}{\mathcal{D}_0(z)^2}(1+z')^2 P_\delta\left(k=\frac{\ell}{\mathcal{D}_0(z')};z'\right). \tag{1.38}$$

I do not detail this small computation since it is very similar to more general results such as equation (3.4) and the derivation of the post-Born correction to the skewness that I do derive in detail in section 1.7, but still note that I use $\Delta\phi = 2\delta_{\rm mass}$ to get to equation (1.38), relying on the fact that the term $\phi_{,33}$, where the direction 3 is along the line-of-sight, cancels out in the integration. This corresponds to a plane-parallel approximation, that is to say, only waves perpendicular to the line-of-sight contribute to lensing, which holds in small-angle cases and/or for slowly varying projection kernels and is again sometimes referred to as the Limber approximation. At the scales of interest here (namely below the degree), this approximation is extremely accurate (see for example Lemos et al. (2017) for an illustrative plot for the power spectrum). I also discuss in somewhat more detail the Limber approximation in the case of the BNT transform in section 1.9.5.

So far, this subsection has established that the measured ellipticities of galaxies give an un-biased estimator of the (reduced) shear. However, as might be already apparent, it is much easier to perform computations at the level of the convergence field, which is a direct projection of the over-densities along the line of sight, being the Laplacian of the projected gravitational potential. It is thus convenient to define the *aperture mass*, a direct observable from the shear field and which can also be easily linked to the convergence. Intuitively one could already sketch the form of this quantity – it is going to be some sort of filtering of the shear field that should also be expressed as another smoothing of the convergence minus a boundary term that illustrates the fact that the convergence can be recovered from the shear only up to a mass-sheet degeneracy. As such, a compensated filter – a filter that averages to zero – on the convergence will do the trick. The aperture mass $M_{\rm ap}$ is thus defined as a

*Actually the mass-sheet degeneracy is a bit more precise than this, since the reconstruction of κ is invariant under the transformation $\kappa \to \kappa' = \lambda\kappa + (1-\lambda)$, λ being an arbitrary constant. This can be seen by transforming the projected gravitational potential $\psi(\boldsymbol{\theta}, z) \to \psi'(\boldsymbol{\theta}, z) = \frac{1-\lambda}{2}\boldsymbol{\theta}^2 + \lambda\psi(\boldsymbol{\theta}, z)$. Then the shear becomes $\gamma(\boldsymbol{\theta}, z) \to \lambda\gamma(\boldsymbol{\theta}, z)$, κ becomes κ' and the reduced shear, the true observable, is left unchanged.

geometrical average of the local convergence with a window of vanishing average

$$M_{\rm ap}(\boldsymbol{\vartheta}) = \int \mathrm{d}^2\boldsymbol{\vartheta}'\, U_\theta(\vartheta')\, \kappa(\boldsymbol{\vartheta}' - \boldsymbol{\vartheta}) \tag{1.39}$$

with

$$\int \mathrm{d}^2\boldsymbol{\vartheta}'\, U_\theta(\vartheta') = 0. \tag{1.40}$$

For its expression as a function of the shear, it is convenient to decompose the two-component shear into its *tangential* γ_t and *cross* $\gamma_\times$ components which, with respect to a given direction vector $\boldsymbol{\theta}$ whose polar angle is ϕ, are defined as

$$\gamma_t = -\Re\left(\gamma \mathrm{e}^{-2\mathrm{i}\phi}\right); \quad \gamma_\times = -\Im\left(\gamma \mathrm{e}^{-2\mathrm{i}\phi}\right). \tag{1.41}$$

This then allows the definition of the aperture mass as (Kaiser, 1995; Schneider, 1996)

$$M_{\rm ap}(\boldsymbol{\vartheta}) = \int \mathrm{d}^2\boldsymbol{\vartheta}' Q_\theta\left(\vartheta'\right) \gamma_t\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\right), \tag{1.42}$$

where

$$Q_\theta(\vartheta) = -U_\theta(\vartheta) + \frac{2}{\vartheta^2}\int_0^\vartheta \mathrm{d}\vartheta'\vartheta' U_\theta\left(\vartheta'\right), \tag{1.43}$$

thus rendering the aperture mass a direct observable up to a reduced shear correction.

1.5.3 Realistic distribution of sources

In realistic scenarios, sources are not confined in discrete source planes but rather continuously distributed in redshift. However, considering a realistic distribution of sources is actually fairly straightforward. Indeed, what one observes is the average of the signal emitted by each source and thus one has to add to the definition (1.30) of A an integration along the sources modulated by the distribution function of the sources $n(z_s)$ along the line of sight:

$$A(z, z_s) \to \int_{\rm sources} \frac{\mathrm{d}z_s}{H(z_s)} n(z_s) A(z, z_s), \tag{1.44}$$

and

$$\int_{\rm sources} \frac{\mathrm{d}z_s}{H(z_s)} n(z_s) = 1. \tag{1.45}$$

In the Euclid survey (one of the stage IV galaxy surveys, another example being LSST) we aim to divide the width of the survey into equally populated source bins (with the same galaxy density in each bin) to get an insight into the clustering of matter at different redshifts (Laureijs et al., 2011). We thus define a global $n(z_s)$ for the survey and as many A matrices as the number of source bins. However, the distribution of sources appearing in equation (1.44) must now be the distribution normalised within this particular bin, with the integration running from the minimum to maximum redshift of the bin. Several reasons push us towards the use of multiple redshift bins. First of all, considering bins of finite size rather than source planes allows us to increase the galaxy density and thus trigger

better measurements. Another reason may also come from errors in the estimation of the redshift, particularly photometric redshifts, which are statistically reduced when considering galaxies within a rather large bin. However, the bigger the number of bins the more sensitive we are to the redshift evolution of the clustering of matter and thus a compromise must be found between the sensitivity of the instrument, our ability to measure redshift, the gain in cosmological information of having many bins, and thus our pretensions in terms of constraints on the parameters we aim at measuring.

Using the continuous distribution of sources along the line of sight is also at the heart of another analysis method of cosmic shear experiments known as 3D lensing (Heavens, 2003). In this framework, one makes full use the redshift information of each individual galaxy instead of binning them into redshift slices, and samples the three-dimensional (reduced-)shear field by performing a spherical harmonics transform of the field into ℓ-modes on the sky, and k-modes along the line of sight. This method is not of concern in this thesis but still needs mentioning as it was popular at least a few years ago, in the decade 2000-2010.

1.5.4 Additional corrections

To continue on the subject of realistic scenarios, I also want to briefly cover the traditional sources of error that limit our ability to precisely estimate the values of our observables. I basically divide them into two categories, additional physical effects and approximations that must be disentangled from what we aim at measuring or accounted for to refine measurements, and real "human" sources of error that come from our instruments and methods of measuring the shapes and positions of galaxies. The former is treated in this subsection and the latter in the next.

Shape noise: Since cosmic shear measurements are performed by measuring the shapes of galaxies, which are themselves intrinsically non-spherical, the observed shear is the sum of contributions from weak-lensing and the intrinsic galactic ellipticities. I have already mentioned that in the absence of intrinsic alignments we might expect the mean value of the intrinsic ellipticity to be zero, but it is not the case for its variance. Shape noise is thus caused by the variance of the intrinsic ellipticity, and is the dominant source of noise in shear measurements, though we seem to be able to model it quite reliably. We model it considering only this variance, that is as Gaussian noise of zero mean and variance σ^2_{SN} and with no spatial correlations. Note that shape noise is important for auto-correlation spectra (which include one-point statistics, that is moments and cumulants), but not for cross-correlations between sources at different comoving distances since their intrinsic ellipticities are thus uncorrelated. To estimate the variance of the shape noise distribution in a Euclid-like configuration, we assume

$$\sigma^2_{SN} = \sigma^2_\epsilon / (n_{g_s} \Omega_\theta), \tag{1.46}$$

where $\sigma^2_\epsilon = 0.3$ is the typical variance of the intrinsic ellipticity of a galaxy, Ω_θ is a solid angle in units of arcmin^2 – typically the area of a pixel – and n_{g_s} is the normalised mean density of observed galaxies in the source plane or source-bin considered (Deshpande et al., 2020).

Intrinsic alignments: Intrinsic alignments (IA) – originating from the correlation between the orientation of galaxies and their environment such as tidal fields – represent a primary contamination for cosmic shear measurements, both because of the amount to which they contribute to the measured shear power spectrum (roughly 10% in loss of power across all angular scales, see figure 8 of Troxel & Ishak (2015) for an example, as well the entire review and references for a complete overview of the modelling of intrinsic alignments), and because they are also quite hard to model. The basic idea is the following. When the observed shape of a galaxy is decomposed into an intrinsic term and a contribution coming from weak lensing along the line-of-sight, two sorts of correlators arise, the auto-correlation spectra of intrinsic shapes (II in the case of the two-point spectrum) and the cross-correlation spectra between foreground intrinsic shapes and background shear (GI still in the case of the two-point spectrum). The GI term exists because some matter creating tidal field located at some redshift z_i is both responsible for part of the alignments of galaxies located at redshift z_i and the part of the lensing effect on some galaxies located along the same line of sight at redshift $z_j > z_i$. However, there is no contribution coming from the correlation of foreground shear and background intrinsic shapes (IG) since generally, the shear-induced ellipticity of a galaxy can only be correlated to objects located within its radial selection function, or more precisely, far away background objects do not contribute to the shape of a galaxy and IA only occur on smaller scales (typically below a hundred megaparsecs, therefore not between distant redshift bins). Nowadays strategies to model this effect rely for example on semi-analytical models describing the two components of the induced shear, which work as a bias expansion in the tidal tensor s_{ij} (which is itself roughly second derivatives of the smoothed Newtonian gravitational potential)

$$\gamma_{ij}^I = \underbrace{C_1 s_{ij}}_{\text{Tidal Alignment}} + \underbrace{C_{1\delta}\left(\delta \times s_{ij}\right)}_{\text{Density Weighting}} + \underbrace{C_2 \left[\sum_{k=0}^{2} s_{ik}s_{kj} - \frac{1}{3}\delta_{ij}s^2\right]}_{\text{Tidal Torquing}} + \dots, \tag{1.47}$$

where C_1 and C_2 are not given by the expansion and represent the amplitude of each contribution. The more precise meaning of those terms will not be given here but can be found in Blazek et al. (2019), which first constructed this type of expansion by gluing together what was known as linear alignments (tidal alignments), density weighting and quadratic alignments (tidal torquing) and taking into account all other possible terms in the expansion. The measurements, or at least constraints from hydro-dynamical numerical simulations (that is simulations that take into account the physics of gas and galaxies additionally to pure gravity, see the next paragraph for baryonic feedback), of the amplitude of those parameters is still an active field of research which suffers from the lack of large volume simulations* (Samuroff et al., 2021).

Baryonic feedback: In all the computations that are carried out throughout this thesis and in the numerical simulations that are used, both dark matter and baryons are assumed

*Big hydro simulations are much more computationally expensive than pure gravity and are therefore of the order 100 Mpc/h, pretty far from cosmological volumes of the order of 1 Gpc/h needed to have enough statistics considering that the typical size of a structure in the Universe, say a void, is 10-100 Mpc/h.

to only interact gravitationally. However, in more realistic scenarios that seriously take baryons into account, the baryonic gas is allowed to cool and to form stars at the centres of dark matter halos. At the same time, feedback effects from active galactic nuclei (AGN, usually black holes) and supernovae may heat up the gas and push it towards the outskirts of halos, and the dark matter profiles might contract or expand reacting to the presence of stars or gas. These processes thus have an impact on the clustering of matter with a maximum suppression of $15-25\%$ around $k\sim 10h/\text{Mpc}$ on the matter power spectrum at $z\leq 1$ and decreasing a bit at higher redshift. This obviously translates to the weak-lensing angular spectrum where we could expect a suppression of 10-20 percent at scales beyond $\ell\sim 200-1000$ (Schneider et al., 2019a; Gouin et al., 2019; Schneider et al., 2019b). However, though there is nowadays a relative convergence on the results of the discretization of Euler equations for the dynamics of the gas, be it through smoothed particles hydrodynamics or finite volume methods, the real issues for (small because hydro) cosmological simulations lies in all the physics that happens below the resolution (sub-grid) of the simulations. These effects concern notably the feedback of supernovae and AGN, and there is nowadays no real consensus on a correct sub-grid model to implement them in simulations. Indeed, though stellar and black hole physics help us understand rather precisely the causes, timing and occurrences of the feedback itself, it is the propagation of the released energy – modelled as a Dirac delta at a given instant in the simulation – that causes some issues, at least for its propagation at scales not resolved in the simulation. The basic idea is then to fit semi-analytical prescriptions coming from ultra-high resolution (sub-parsec) simulations designed to directly experience the feedback, but different, although motivated, recipes implemented in different simulations unfortunately have different and noticeable impacts on the physics of galaxies (the star formation rate for example) as well as on the clustering of matter at larger scales. This may have to do with the sensitivity of the feedback energy propagation to its (non-homogeneous, porous) environment, which is still very difficult to take into account. A rather complete review of different sub-grid models and the issues I presented here can be found in Naab & Ostriker (2017). Nevertheless, this means that accurate cosmological inference for our field must be done in regimes relatively safe from baryonic feedback and as an illustration, Weiss et al. (2019) for instance suggested a cut at 16 arcmin for a Euclid-like survey in order to safely ignore the physics of baryons. We will see that this is yet another motivation for the BNT transform proposed in section 1.9.

Reduced shear correction: Again neglecting the intrinsic alignments of galaxies, I previously showed that the measured ellipticites of galaxies are an unbiased estimator of the reduced shear $g=\gamma/(1-\kappa)$ rather than the shear itself. Expanding g around $\kappa=0$, the leading order correction thus consists in replacing the shear by an observed shear $\gamma_{\text{obs}}=\gamma+\gamma\kappa$ in every correlator computation taking advantage of the relation (1.36) between γ and κ. I will provide an example of this type of calculation when computing this correction for the aperture mass skewness in section 1.9.6 in the context of the BNT transform. The effect itself is however relatively small at least on the shear power spectrum as it is shown to be a correction of less than one percent across all scales and redshifts probed by a Euclid-like experiment (see figure 1 of Deshpande et al., 2020). It is, however, often combined with another correction, the magnification bias, which has a similar functional form and can be

more important especially at small scales and high redshifts in a Euclid-like experiment ($\ell \sim 1000$ and $z \sim 2$).

MAGNIFICATION BIAS: As a consequence of the magnification effect discussed in section 1.5.6, faint sources at the flux limit of a survey might be included in the observed sample while they would, in the absence of lensing, be excluded. At the same time, this magnification effect not only changes the object sizes, but also stretches entire regions of the sky behind lenses which thus reduces the number density of source galaxies in those regions. Hence those two effects are in competition and the net result depends on the slope of the intrinsic, unlensed, galaxy luminosity function at the survey's flux limit. This is known as magnification bias. Additionally, galaxies could also be pulled into a sample because their effective radius is increased as a consequence of magnification, such that they pass a resolution factor cut, but this effect is generally not considered for space-based experiments like Euclid, and I will thus also ignore it. Now following the prescription given in Deshpande et al. (2020) that was derived in Hui et al. (2007), the resulting "observed" shear (and not the reduced shear since that would lead to sub-dominant higher couplings) reads

$$\gamma_{\rm obs} = \gamma + \gamma\delta^{\rm g} + (5s - 2)\gamma\kappa, \tag{1.48}$$

where $\delta^{\rm g}$ is the intrinsic, unlensed, galaxy overdensity at the source (or in the redshift bin),

$$s = \left.\frac{\partial \log_{10}(n(z_s, m))}{\partial m}\right|_{m_{\rm lim}}, \tag{1.49}$$

$n(z_s, m)$ is the true distribution of galaxies evaluated at the source (central redshift of the bin) and at a given magnitude (luminous flux) m, and $m_{\rm lim}$ is the survey's limiting magnitude. Note that up to a source-lens clustering term $\gamma\delta^{\rm g}$, the correction has the same form as the reduced shear correction, which often leads to the two being treated together. Indeed, the $\gamma\delta^{\rm g}$ term will yield a zero contribution in any correlator if the lenses and sources do not overlap. This is the case for auto-correlators when considering source planes or sufficiently narrow redshift bins, but more importantly this is also the case when applying the BNT transform studied in section 1.9 to any tomographic bins. Still looking at figure 1 of Deshpande et al. (2020), we observe that the combination of the reduced shear and magnification bias correction for the shear power spectrum of a Euclid-like experiment yields a correction that can be quite important (a few percent) at small scales and high redshifts in a Euclid-like experiment ($\ell \sim 1000$ and $z \sim 2$).

REDSHIFT-SPACE DISTORTIONS: Redshift-space distortions (RSDs) refer to closely related effects that influence the true redshifts of galaxies as opposed to the "cosmic" redshifts caused by the expansion of the universe (Hubble flow). These can both be seen as contamination for lensing experiments since they affect our ability to measure cosmic redshift, which we use as a measure of the radial distance and as sources of cosmological information in their own right (Percival & White, 2009) or tests of gravity on large-scales (Raccanelli et al., 2013). Three effects, of which the first two are physically the same, are usually retained. The *Finger-of-God effect* is where the galaxy distribution is elongated in redshift space,

with an axis of elongation pointed toward the observer (Jackson, 1972). It is caused by a Doppler shift associated with the random peculiar velocities of galaxies bound in structures such as clusters. The large velocities that lead to this effect are associated with the gravity of the cluster by means of the virial theorem and thus change the observed redshifts of the galaxies in the cluster. The exact same physical effect but occurring at larger scales is known as the *Kaiser effect*, where the distortion is caused by the coherent motions of galaxies that coherently fall towards the cluster centre as it assembles (Kaiser, 1987). The Kaiser effect usually leads not to an elongation but to an apparent flattening ("pancakes of God") of the structure. It is a much smaller effect than the Finger-of-God effect, and occurs as mentioned on larger scales. The modelling of RSDs thus takes into account a linear term that acts on large scales (the Kaiser effect, Kaiser (1987)), which comes from SPT, and a non-linear prescription (Finger-of-God) for which a perturbative treatment is not possible and phenomenological modelling is needed (Scoccimarro et al., 1999). This gives rise to an anisotropic galaxy power spectrum $P(k,\mu)$, which can be written from the isotropic matter power spectrum $P(k)$ as

$$P(k,\mu) = \widetilde{A}^2_{\rm RSD}(\mu, k\mu)P(k), \tag{1.50}$$

where k is the radial part of the 3D wave-vector, $\mu = \cos\theta$ where θ is the angle between the wave-vector and the line of sight, and $\widetilde{A}_{\rm RSD}$ is the RSD operator, which takes the power spectrum into redshift space and accounts for galaxy bias. We typically have

$$\widetilde{A}_{\rm RSD}(\mu, k\mu) = b_g\left(1+\beta\mu^2\right)\widetilde{A}_{\rm nl}(\mu, k\mu) \tag{1.51}$$

where $b_g\left(1+\beta\mu^2\right)$ is the Kaiser term with $\beta = f/b_g$, b_g being the linear galaxy bias ($\delta_{\rm galaxies} \simeq b_g\delta_{\rm matter}$) and $f \equiv \mathrm{d}\log D/\mathrm{d}\log a$ the logarithmic linear growth factor given in equation (3.58). The non-linear RSD model is then accounted for by the operator $\widetilde{A}_{\rm nl}$, which is typically phenomenologically chosen to be Gaussian or Lorentzian with variance σ_v^2, the latter being as accurate as more physically motivated models (Jalilvand et al., 2020)

$$\begin{aligned}\widetilde{A}_{\rm nl}(\mu, k\mu) &= \exp\left[-f^2\sigma_v^2k^2\mu^2\right] \quad \text{or}\\ \widetilde{A}_{\rm nl}(\mu, k\mu) &= \frac{1}{1+f^2\sigma_v^2k^2\mu^2}.\end{aligned} \tag{1.52}$$

Finally, the third effect (or second depending on if one considers the Kaiser and Finger-of-God effects to be the same), much less relevant in the content of cosmic shear experiments and more relevant to studies of the anisotropies of the CMB (Crittenden & Turok, 1996), is due to photons coming from distant galaxies traversing potential wells and voids but not fully gaining or losing the appropriate amount of energy when coming out of them. This can be explained by the fact that the accelerated expansion of the universe caused by dark energy causes the decay of even large potential wells and voids over the time it takes a photon to travel through them. This known as the *integrated Sachs-Wolfe* (Sachs & Wolfe, 1967) effect when the photons come from the last scattering surface (at the redshift of the CMB).

Source-source and source-lens clustering: Every derivation that I have presented (and that I will present) in this thesis ignores the effect of the clustering of the galaxies themselves. This clustering gives rise to two separate effect that influence the power spectrum

and skewness at the few percent level (Hamana et al., 2002; Schneider et al., 2002). The source-source clustering takes into account that the correlation of the shear of galaxies at different but close redshifts depends on their angular separation (exactly because of clustering), which tends to generate cross-components of the shear. As for source-lens clustering, the effect exists only for broad distributions of source galaxies, which tend to be more and more rare with the improvement of instruments. Imagine two lines of sight towards a broad distribution of sources. The first passes through a strong gravitational potential at intermediate range and thus strongly lenses the background galaxies. But because source galaxies somehow trace the matter field, there are also sources bound to this gravitational potential that are thus less lensed (because the lens is closer to them and they are within the lens). On average for this line of sight, the foreground sources tend to counteract the high κ that we would measure only considering the background galaxies. Now imagine a second line of sight that passes through a void and reaches background galaxies. These galaxies will generate a strongly negative κ that will not be compensated by foreground galaxies. Thus on average, and for the two lines of sight, the source-lens clustering effect will tend to push the κ distribution towards more negative values than what it would have in the absence of the effect. Note however that though these effects have a small impact on the measured shear in a given source bin, they are themselves probes of the cosmological parameters and often combined with shear measurements in the well-known 3×2 point analysis, which I will present in the introduction of chapter 3.

Additional corrections that could be mentioned also include the so called post-Born corrections, to which I dedicate the two following sections to provide a complete derivation.

1.5.5 Sources of shear observational error

Point spread function: The point spread function (PSF) is the response of the image system (the CCD detectors in a telescope) to a point source. The intrinsic size of this PSF is extremely important since it will tend to circularise galaxy images, thus resulting in a decrease in their shape correlations. This effect typically comes from the atmosphere if applicable, optical aberrations, mirror deformations, tracking errors, CCD non-flatness and misalignments on the focal plane, and pixelation. Note that the typical PSF effect in ground-based images is up to 10%, which means that for a sub-percent measurement of a 1% effect that is the cosmic shear, the PSF induced on galaxy images would have to be corrected with a precision of one part in 100. This is quite a challenge! Thus to estimate and deconvolve the PSF from a galaxy image, one has to select stars on the image, measure their shape, and interpolate the resulting PSF to the position of the galaxy. Indeed, stars are unresolved objects on the image and thus a local representation of the PSF. This requires a suitable sample of stars, *i.e.* in regions not polluted by saturated pixels or cosmic rays and uncontaminated by galaxies, and it also requires a good measurement of the shape of those stars in the image. Note that things are also made complicated by the wavelength dependence of the PSF and the interpolation of the PSF model to the galaxy position, which is not straightforward and influences the values of the measured cosmic shear (Kilbinger, 2015).

Photometric redshift: The sensitivity of most cosmic shear observables to the distribution in redshift of the sources is quite important (Kilbinger, 2015) and thus the importance of determining rather precisely the redshift of all sources. However, performing the spectroscopy of all the faint galaxies used in a typical weak-lensing survey is too costly, and thus redshifts have to be estimated from broad-band photometry using the technique of photometric redshifts, sometimes called *photo-z*. There are many methods on the market to estimate photo-z, but one of the most popular relies on fitting redshifted templates of the spectral energy distributions (SEDs) of galaxies to the flux observed in different bands. For example, the French code *LePhare* (Ilbert et al., 2006), with which I was lucky to work during my first master's internship, applies this method. Fortunately, the methods to estimate photo-z are also usually able to provide error bars on the estimate and even provide the full probability distribution. It is often a redshift-dependent Gaussian of dispersion $\sigma/(1+z) = 0.03-0.06$ (0.05 in Euclid Collaboration et al. (2020) for the Euclid experiment) and the rate of catastrophic outliers – galaxies whose estimated redshift is off from the true (spectroscopic) redshift by more than a couple of standard deviations – is between a few and a few tens of percent (10% in Euclid Collaboration et al. (2020) for the Euclid experiment).

1.5.6 Magnification: luminous flux of observed galaxies

Neglecting the scattering and absorption of observed photons in the late universe, the Boltzmann equation for the photon distribution function has no collision term and the distribution function is thus conserved. As such, the specific intensity* measured in any experiment only changes through the change in frequency of photons due to cosmological/gravitational redshifts and the Doppler effect. This in particular means that the specific intensity with or without gravitational lensing received by the observer is the same, with the only difference being that the source angular position on the sky might not be the same. In other words, the surface brightness of objects is conserved. However, since the apparent size of resolved background objects change through lensing, then their total flux changes as well. This effect is known as *gravitational magnification* and can be used as a weak-lensing observable complementary to the measurement of galaxy shapes.

The magnification μ is formally defined as the ratio between the lensed and the unlensed flux, and since the flux is the integration of the surface brightness over the area of the source, then the magnification is given by the ratio between the lensed and the unlensed area. As a consequence we have

$$\mu = \det A^{-1} = \frac{1}{(1-\kappa)^2 - |\gamma|^2} \approx 1 + 2\kappa, \tag{1.53}$$

where the last right hand side is valid in the weak lensing approximation. Magnification is thus another observable of weak-lensing experiments that this time probes the convergence rather than the shear. Furthermore, magnification comes for free in any weak-lensing survey since fluxes have to be measured with high accuracy anyway to determine photometric redshifts. The requirements for image quality are similar for measuring shapes, sizes or fluxes. Of course, as for any observational method, there are a number of systematics that render the measurements of the cosmic magnification more difficult than they might first appear. A short but good overview is given in Kilbinger (2015) and references therein and

*The energy incident on a detector per solid angle, per unit area, per time and per unit frequency.

I will only summarise them. The signal-to-noise ratio of magnification is slightly worse than that for cosmic shear, and the intrinsic distribution of the size estimator $\sigma_{\log r/r_0} \sim 0.4$ (galaxies have an intrinsic size just as they have an intrinsic ellipticity, which averages to zero but with a non-zero variance) is also slightly bigger than the ellipticity estimator $\sigma_\epsilon \sim 0.3$ (see section 1.5.4). There also exist the analogue of intrinsic alignments, which are intrinsic magnitude and size correlations, and which are caused by the clustering of galaxies (the average of the size and magnitude estimators no longer average to zero). Intrinsic magnitude correlations are about a factor of 10 larger than intrinsic alignments, while intrinsic size correlations are smaller. Note however, that this last point can be mitigated when correlating background to foreground galaxies. Moreover, flux measurements can be performed on galaxies that are too small or faint for accurate shape measurements, which makes them complementary to cosmic shear. I will not insist anymore on magnification as it is not the primary focus of this thesis, which rather deals with cosmic shear experiments, but it is still nevertheless interesting to mention that the convergence is also an observable in itself, at least at the same level as the shear, within the weak-lensing approximation.

1.6. Convergence at second order

The aim of this section is to compute all second-order corrections coming from our derivation of the deformation matrix – so called post-Born corrections – on the definition of the convergence*. This is indeed necessary to compute their impact on the third one-point cumulant of the field (skewness) in the next section, the computation being of particular relevance in the context of the one-point probability distribution function which I studied a lot during the course of my thesis. More generally, this serves as a good illustration of the fact that these corrections are not significant in most modern cosmic shear surveys and is a good exercise/example of the type of computations that are done using SPT.

Let me recap a bit. The convergence field is defined as the isotropic deformation of a bundle of light and is thus very closely related to the trace of the deformation matrix. It is defined as

$$\kappa = 1 - \frac{1}{2}\mathrm{Tr}\left[\frac{\mathcal{D}_{ab}}{\mathcal{D}_0}\right]. \tag{1.54}$$

From equations (1.27) and (1.29) that define the deformation matrix at first and second order, we get the convergence at first and second order (the zeroth order is obviously null from the definition (1.54))

$$\kappa^{(1)}(\boldsymbol{\theta}, z) = \frac{3}{2}\Omega_m \int_0^z \frac{\mathrm{d}z'}{H(z')} \frac{\mathcal{D}_0(z', z)\mathcal{D}_0(z')}{\mathcal{D}_0(z)} (1+z')\delta^{(1)}_{\mathrm{mass}}(\boldsymbol{\theta}, z'), \tag{1.55}$$

$$\kappa^{(2)}(\boldsymbol{\theta}, z) = \frac{3}{2}\Omega_m \int_0^z \frac{\mathrm{d}z'}{H(z')} \frac{\mathcal{D}_0(z', z)\mathcal{D}_0(z')}{\mathcal{D}_0(z)} (1+z')\delta^{(2)}_{\mathrm{mass}}(\boldsymbol{\theta}, z') + \kappa^{(2)}_{\mathrm{corr.1}}(\boldsymbol{\theta}, z) \tag{1.56}$$

*Technically other terms would appear in a full-sky relativistic setup but they would only matter at very large angular scales while we are rather interested in the non-linear dynamics that occur at relatively small scales, typically a few to a few tens of Mpc/h (Bernardeau et al., 2010)

where $\kappa^{(2)}_{\rm corr.1}(\boldsymbol{\theta}, z)$ gathers contributions from the coupling of lenses that are second order in the field (there is none at first order as stated above for the deformation matrix) and will be explicitly given below. Once again, I used $\Delta\phi^{(n)} = 2\delta^{(n)}_{\rm mass}$ to get equations (1.55) and (1.56) relying on the fact that the term $\phi_{,33}$, where the direction 3 is along the line-of-sight, cancels out in the integration (Limber approximation).

Let me also define the weighting factor along the line-of-sight

$$\omega(z', z) = \frac{3}{2}\Omega_m \frac{\mathcal{D}_0(z', z)\mathcal{D}_0(z')}{\mathcal{D}_0(z)}(1 + z'), \tag{1.57}$$

also called lensing efficiency function or lensing kernel, and which characterises the efficiency of the lenses along the line-of-sight. This simplifies the definition and interpretation of κ and might also be abridged in $\omega(z')$ when the redshift z of the source is obvious.

The convergence at second order comprises three different terms that I will now describe in more detail:

Born approximation with independent lenses: First, as shown in the previous section, the convergence κ gets a contribution at every order from the integration along the line of sight of the same order density contrast. This contribution tends to be the dominant term. Thus it is common to approximate the convergence by the density field projected along the line of sight such that

$$\kappa(\boldsymbol{\theta}, z) = \int_0^z \frac{{\rm d}z'}{H(z')}\omega(z', z)\delta_{\rm mass}(\boldsymbol{\theta}, z'), \tag{1.58}$$

where $\delta_{\rm mass}(\boldsymbol{\theta}, z')$ is the non-linear matter density contrast. However, we have seen that this approximation is strictly valid only in the linear regime and some corrections arise at second order and beyond. Thus, only in this approximation is the convergence *potential*, which means it is seen as the Laplacian of some projected gravitational potential and thus "easily" expressed in terms of the shear as already shown in the previous section 1.5.

Lens-lens coupling: At second order, the correction accounts for the couplings of lenses along the line-of-sight and can be written as*

$$\kappa^{(2)}_{\rm corr.1}(\boldsymbol{\theta}, z) = -\frac{1}{2}\int_0^z \frac{{\rm d}z'}{H(z')}\omega(z', z)\int_0^{z'} \frac{{\rm d}z''}{H(z'')}\omega(z'', z')\phi^{(1)}_{,ab}(\boldsymbol{\theta}, z')\phi^{(1)}_{,ab}(\boldsymbol{\theta}, z'') \tag{1.59}$$

which comes directly from equations (1.29) and (1.56). This term is exactly the same as the second term of equation (1.29) already discussed, but now expressed for the convergence rather than the full deformation matrix. Being typically negative, this corrective term then contributes to Gaussianising the field and lessens the importance of quantities beyond the two-point correlation function. In fact, all second order coupling terms will tend to Gaussianise the convergence field, since they characterise the introduction of more random deflections along the light path, which will in turn tend to diminish the impact of the non-linear clustering of matter. I find that a good mental image to have is that of clustered chunks of matter blurred by these lensing terms and thus appearing less compact.

*Let me re-emphasize here that there is an implicit summation over the repeated indices and that indices a and b run only on the transverse directions due to the use of the Limber approximation.

Geodesic deviation: In the derivation of the deformation matrix and thus the convergence, I made use of the Born approximation and therefore integrated along the unperturbed line of sight. However, this approximation is not valid in a lumpy universe and one needs to account for additional couplings between lenses due to the perturbations of the trajectories of photons. Hence, let me now include in the line-of-sight integrations the fact that the potential and density are not to be taken at angular position $\boldsymbol{\theta}$, as in equation (1.55), but at $\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta}$ where $\mathrm{d}\boldsymbol{\theta}$ is the deflection induced by the foreground lenses. To compute its value, first remember that $A_{ij} = \mathcal{D}_{ij}/\mathcal{D}_0$ is the Jacobian of the local transformation from the real (source) angle to the apparent (image) angle. Then considering a lens equation of the type $\theta_i^{\mathrm{O}} = \theta_i^{\mathrm{S}} + \mathrm{d}\theta_i\left(\theta^{\mathrm{O}}\right)$ provides us with a direct link between the deflection angle $\mathrm{d}\theta_i\left(\theta^{\mathrm{S}}\right)$ we are looking for and $\mathcal{D}_{ij}$ since we find that

$$A_{ij} = \delta_{ij}^D + \frac{\partial \mathrm{d}\theta_i\left(\theta^{\mathrm{O}}\right)}{\partial \theta_j^{\mathrm{O}}}. \tag{1.60}$$

Then using equation (1.27) and being careful that derivatives are taken with respect to transverse spatial components and not angular, one gets at first order

$$\mathrm{d}\theta_a^{(1)}(\boldsymbol{\theta}, z) = -\int_0^z \frac{\mathrm{d}z'}{H(z')} \frac{\omega(z', z)}{\mathcal{D}_0(z')} \phi_{,a}^{(1)}(\boldsymbol{\theta}, z'). \tag{1.61}$$

Now Taylor expanding the density contrast

$$\delta_{\mathrm{mass}}^{(1)}(\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta}, z) \approx \delta_{\mathrm{mass}}^{(1)}(\boldsymbol{\theta}, z) + \nabla_{\boldsymbol{\theta}} \delta_{\mathrm{mass}}^{(1)}(\boldsymbol{\theta}, z) \cdot \mathrm{d}\boldsymbol{\theta}, \tag{1.62}$$

and combining equations (1.62) and (1.55) yields an expression for the leading order correction to the Born approximation which is second order for the convergence

$$\kappa_{\mathrm{corr.2}}^{(2)}(\boldsymbol{\theta}, z) = -\frac{1}{2} \int_0^z \frac{\mathrm{d}z'}{H(z')} \omega(z', z) \int_0^{z'} \frac{\mathrm{d}z''}{H(z'')} \omega(z'', z') \frac{\mathcal{D}_0(z')}{\mathcal{D}_0(z'')} \phi_{,aab}^{(1)}(\boldsymbol{\theta}, z') \phi_{,b}^{(1)}(\boldsymbol{\theta}, z''), \tag{1.63}$$

where derivatives are still taken with respect to transverse spatial components. Note that similarly to (1.59), this correction also tends to Gaussianise the field. Finally note that all the derivations of lensing quantities we presented here are to be found in many books and references using the so-called Fermat's principle. This approach leads to the same equations and is thus equivalent to our formalism and I propose a quick derivation/summary in appendix A.2.

1.7. Post-Born corrections to the skewness

The computation of post-Born corrections to the convergence skewness serves here two purposes: first, it is a very good example of the kind of computation that is performed in the development of theoretical weak-lensing tools but with the advantage that this particular derivation is not present in every textbook on the subject, and second, it is also a good example of the kind of justification of the Born-and-independent-lenses approximation that is often used. In the literature, it is a rather well-known fact that post-Born corrections are only relevant for sources at very high redshift close to the last scattering surface – typically CMB lensing – and especially for observables beyond the power spectrum, such as the bi-spectrum or PDF. Let me first perform a small review of this literature.

1.7.1 Post-Born corrections in the literature

At the level of cosmic shear experiments, that is, for source galaxy redshifts of order 1 and sufficiently large scales, the power spectrum of the convergence field is very insensitive to post-Born corrections, which, as shown by Shapiro & Cooray (2006), are several orders of magnitudes below the signal, and only become more important than cosmic variance for a full-sky survey at really small scales down to which other uncertainties arise, such as baryonic physics. This was confirmed for instance by Hilbert et al. (2009) with the help of ray-tracing through N-body simulations, which showed that the first-order (in the Newtonian potential) approximation provides us with an excellent fit to cosmic-shear power spectra as long as the exact (fully non-linear) matter power spectrum is used as an input. Using the same simulations, cosmic-shear B-modes, which are induced by post-Born corrections and lens-lens coupling, were also found to be at least three orders of magnitude smaller than cosmic-shear E-modes. Going beyond the convergence and shear, Schäfer et al. (2012) focused on the cosmic flexion (derivatives of the deformation matrix) and showed that post-Born corrections to the flexion power spectra are roughly four order of magnitudes below the signal at redshift of order 1, again much below the cosmic-variance limit. As for non-Gaussian statistics, Bernardeau et al. (1997) for the convergence and Schneider et al. (1998) for the aperture mass both showed that the correction induced on the skewness was also negligible for sources at redshift around 1, the corrections being smaller than the errors induced by using the tree-order perturbation theory result to get the skewness, which was also confirmed extracting those skewness corrections from numerical simulations (Petri et al., 2017). Hence making use of the standard definition of weak-lensing quantities neglecting post-Born corrections is supposedly enough for the majority of weak-lensing experiments as long as very small scales are not considered. Also note that the addition of baryons does not change this picture (Gouin et al., 2019).

In the context of CMB lensing, post-Born corrections also have a negligible effect on power spectra: they appear to be irrelevant for the E-mode temperature and polarisation power spectra and differences in the B-modes power spectra only show up at very small scales ($\ell \geq 3000$) (Pratten & Lewis, 2016; Marozzi et al., 2018; Fabbian et al., 2018). Similar findings are reported on the cross-correlation between the thermal Sunyaev-Zeldovich effect and the observed convergence (with supplementary terms coming from the reduced shear), the correction being several orders of magnitude below the first-order result and under the cosmic-variance limit up until very large angular modes ($\ell \sim 4000$) as shown by Tröster & Van Waerbeke (2014). However, post-Born effects might be important for other CMB lensing observables, for instance when cross-correlating galaxy counts and CMB lensing convergence (Böhm et al., 2019; Fabbian et al., 2019) or when cosmic shear maps are used to reconstruct the lensing potential power spectra for which a significant bias appears when neglecting both post-Born terms and the non-Gaussianity of the large-scale structure (Böhm et al., 2016, 2018; Beck et al., 2018). Post-Born corrections also seem important for higher-order statistics of CMB lensing. As an example, their contribution to the CMB convergence bispectrum is at a comparable amplitude to the signal coming from gravitational non-linearities (Pratten & Lewis, 2016; Fabbian et al., 2018) and therefore has to be accounted for if the bispectrum is used to constrain possible modifications of gravity (Namikawa et al., 2018) although this might not be observable even with stage-IV type experiments (Namikawa et al., 2019).

1.7.2 Derivation

Given the second-order corrections to the convergence obtained in the previous section 1.6, let me now compute the corrections induced on the skewness of the convergence defined as

$$S_{3,\kappa} = \frac{\langle \kappa^3 \rangle_c}{\langle \kappa^2 \rangle_c^2} = \frac{\langle \kappa^3 \rangle}{\langle \kappa^2 \rangle^2}, \tag{1.64}$$

with $\langle . \rangle$ denoting an ensemble average, and since moments and cumulants (connected part of the moments, denoted with the subscript $_c$) are equal at second and third order for a random variable with zero mean $\langle \kappa \rangle = 0$*. The leading order term to the third moment is given by

$$\begin{aligned} \left\langle \kappa^3 \right\rangle &= \left\langle \left(\kappa^{(1)} + \kappa^{(2)} + \dots\right)^3 \right\rangle \\ &= \left\langle \left(\kappa^{(1)}\right)^3 \right\rangle + 3 \left\langle \left(\kappa^{(1)}\right)^2 \kappa^{(2)} \right\rangle + \dots \\ &= 3 \left\langle \left(\kappa^{(1)}\right)^2 \kappa^{(2)} \right\rangle + \dots \end{aligned} \tag{1.65}$$

because the linear convergence does not admit a skewness term if no primordial non-Gaussianities are assumed (as I do in this thesis). As such, the leading order correction to the skewness induced by the second order post-Born corrective terms is given by

$$S_{3,\kappa}^{\rm corr} = 3 \frac{\left\langle \left(\kappa^{(1)}\right)^2 \kappa_{\rm corr}^{(2)} \right\rangle}{\left\langle \left(\kappa^{(1)}\right)^2 \right\rangle^2}. \tag{1.66}$$

Finally, the convergence field itself is not accessible and only a version smoothed by some window-function is of interest. For simplicity, and because much of the following work of this thesis will do the same, I use an angular top-hat window of radius θ which reads in Fourier space $W\left[k_\perp \mathcal{D}_0(z)\theta\right]$ where $k_\perp$ is the norm of $\mathbf{k}$ in the transverse directions and

$$W(l) = 2\frac{J_1(l)}{l}, \tag{1.67}$$

with J_1 the first order Bessel function of the first kind. Note that the linear variance of the convergence field filtered with an opening angle of θ which enters equation (1.66) can be directly computed from equation (1.55) using the Limber approximation. However, since I will later on perform the general computation for every cumulant in the demonstration of equation (3.4), let us just admit for now that

$$\left\langle \left(\kappa_\theta^{(1)}\right)^2 \right\rangle = \frac{1}{2\pi} \int_0^z \frac{{\rm d}z'}{H(z')} \omega(z', z)^2 D_+(z')^2 \int_0^{+\infty} {\rm d}k_\perp k_\perp P_\delta^l(k_\perp) W(k_\perp \mathcal{D}_0(z')\theta)^2, \tag{1.68}$$

where $D_+(z)$ is the linear growth factor and $P_\delta^l(k)$ the linear matter power spectrum.

*The mean convergence is zero by definition since the mean of the density contrast is also zero by definition.

In order to explicitly compute the post-Born corrections to the skewness, let me first express the gravitational potential ϕ in Fourier space as

$$\phi(\boldsymbol{\theta}, z) = \int \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^{3/2}} \tilde{\phi}(\boldsymbol{k}, z) e^{\mathrm{i}(\boldsymbol{k}_\perp \cdot \boldsymbol{\theta} + k_r \chi(z))}, \tag{1.69}$$

which can be related to the density field via a Fourier transform of the Poisson equation (1.20) so that at first order

$$\tilde{\phi}^{(1)}(\boldsymbol{k}, z) = \frac{2}{k^2} D_+(z) \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}). \tag{1.70}$$

From there, one can easily express the terms $\phi^{(1)}_{,ab}(\boldsymbol{\theta}, z)\phi^{(1)}_{,ab}(\boldsymbol{\theta}, z')$ and $\phi^{(1)}_{,aab}(\boldsymbol{\theta}, z)\phi^{(1)}_{,b}(\boldsymbol{\theta}, z')$ entering the post-Born corrections in equations (1.59) and (1.63). First, the contraction of second derivatives can be written

$$\begin{aligned}\phi^{(1)}_{,ab}(\boldsymbol{\theta}, z)\phi^{(1)}_{,ab}(\boldsymbol{\theta}, z') = \frac{4}{(2\pi)^3}\int \mathrm{d}\boldsymbol{k}\mathrm{d}\boldsymbol{k}' \left[\frac{(\boldsymbol{k}\cdot\boldsymbol{k}')^2}{k^2 k'^2}\right] e^{\mathrm{i}(\boldsymbol{k}_\perp + \boldsymbol{k}'_\perp)\cdot\boldsymbol{\theta}} \\ \times e^{\mathrm{i}(k_r\chi(z) + k'_r\chi(z'))} D_+(z) D_+(z') \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}) \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}'),\end{aligned} \tag{1.71}$$

where the small-angle (Limber) approximation $\boldsymbol{k}^2_\perp \sim \boldsymbol{k}^2$ was used. Similarly, the contraction of third and first derivatives in the geodesic deviation reads

$$\begin{aligned}\phi^{(1)}_{,aab}(\boldsymbol{\theta}, z)\phi^{(1)}_{,b}(\boldsymbol{\theta}, z') = \frac{4}{(2\pi)^3}\int \mathrm{d}\boldsymbol{k}\mathrm{d}\boldsymbol{k}' \left[\frac{\boldsymbol{k}\cdot\boldsymbol{k}'}{k'^2}\right] e^{\mathrm{i}(\boldsymbol{k}_\perp + \boldsymbol{k}'_\perp)\cdot\boldsymbol{\theta}} \\ e^{\mathrm{i}(k_r\chi(z) + k'_r\chi(z'))} D_+(z) D_+(z') \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}) \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}').\end{aligned} \tag{1.72}$$

Equations (1.71) and (1.72) have the same structure. I thus gather them using the kernel

$$G(\boldsymbol{k}, \boldsymbol{k}', z', z'') = \frac{\mathcal{D}_0(z')}{\mathcal{D}_0(z'')} \frac{\boldsymbol{k}\cdot\boldsymbol{k}'}{k'^2} + \frac{(\boldsymbol{k}\cdot\boldsymbol{k}')^2}{k^2 k'^2}, \tag{1.73}$$

such that the corrective terms defined in equations (1.59) and (1.63) can be jointly written as

$$\begin{aligned}\kappa^{(2)}_{\mathrm{corr}}(\boldsymbol{\theta}, z) = -\frac{1}{2}\frac{4}{(2\pi)^3}\int_0^z \frac{\mathrm{d}z'}{H(z')}\omega(z', z)\int_0^{z'} \frac{\mathrm{d}z''}{H(z'')}\omega(z'', z')\int \mathrm{d}\boldsymbol{k}\mathrm{d}\boldsymbol{k}' G(\boldsymbol{k}, \boldsymbol{k}', z', z'') \\ \times e^{\mathrm{i}((\boldsymbol{k}_\perp + \boldsymbol{k}'_\perp)\cdot\boldsymbol{\theta} + k_r\chi(z) + k'_r\chi(z'))} D_+(z') D_+(z'') \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}) \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}').\end{aligned} \tag{1.74}$$

From there, the computation of the post-Born correction to the third order moment is very similar to that of the usual skewness in SPT. I start by recalling Wick's theorem for Gaussian random fields, as is the case for the matter density field at linear order $\delta^{(1)}_{\mathrm{mass}}$:

$$\begin{aligned}\left\langle \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}_1)...\tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}_{2p+1}) \right\rangle &= 0 \\ \left\langle \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}_1)...\tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}_{2p}) \right\rangle &= \sum_{\text{pairs}} \prod_{p \text{ pairs } (i\neq j)} \left\langle \tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}_i)\tilde{\delta}^{(1)}_{\mathrm{mass}}(\boldsymbol{k}_j) \right\rangle.\end{aligned} \tag{1.75}$$

Furthermore, because of the homogeneity and isotropy of our universe

$$\left\langle \tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_1)\tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_2)\right\rangle = \delta^D(\boldsymbol{k}_1+\boldsymbol{k}_2)P^l_\delta(k_1). \tag{1.76}$$

I then make use of Wick's theorem to simplify the integrand appearing when plugging equation (1.74) into (1.66)

$$\begin{aligned}\int \mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2\mathrm{d}\boldsymbol{k}_3\mathrm{d}\boldsymbol{k}_4 \left\langle \tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_1)\tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_2)\tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_3)\tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_4)\right\rangle G(\boldsymbol{k}_3,\boldsymbol{k}_4,z'',z''')e^{\mathrm{i}\boldsymbol{k}_3\cdot\boldsymbol{r}''(z'')}\\ \times e^{\mathrm{i}\boldsymbol{k}_4\cdot\boldsymbol{r}''(z''')}e^{\mathrm{i}\boldsymbol{k}_1\cdot\boldsymbol{r}(z)}e^{\mathrm{i}\boldsymbol{k}_2\cdot\boldsymbol{r}'(z')} = \int \mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2 P^l_\delta(k_1)P^l_\delta(k_2)G(\boldsymbol{k}_1,\boldsymbol{k}_2,z'',z''')e^{\mathrm{i}\boldsymbol{k}_1\cdot\boldsymbol{r}''(z'')}e^{\mathrm{i}\boldsymbol{k}_2\cdot\boldsymbol{r}''(z''')}\\ \times\left[e^{-\mathrm{i}\boldsymbol{k}_1\cdot\boldsymbol{r}(z)}e^{-\mathrm{i}\boldsymbol{k}_2\cdot\boldsymbol{r}'(z')}+e^{-\mathrm{i}\boldsymbol{k}_1\cdot\boldsymbol{r}'(z')}e^{-\mathrm{i}\boldsymbol{k}_2\cdot\boldsymbol{r}(z)}\right]+\int \mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2 P^l_\delta(k_1)P^l_\delta(k_2)\left(1-\frac{\mathcal{D}_0(z'')}{\mathcal{D}_0(z''')}\right)\\ \times e^{\mathrm{i}(\boldsymbol{k}_1\cdot\boldsymbol{r}''(z'')-\boldsymbol{k}_1\cdot\boldsymbol{r}''(z'''))}e^{\mathrm{i}(\boldsymbol{k}_2\cdot\boldsymbol{r}(z)-\boldsymbol{k}_2\cdot\boldsymbol{r}'(z'))},\end{aligned} \tag{1.77}$$

where $\boldsymbol{r}(z)=(\chi(z),\boldsymbol{\theta})$. Then filtering the field with an opening angle of θ and integrating over the radial wave vectors we get

$$\begin{aligned}\int \mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2\mathrm{d}\boldsymbol{k}_3\mathrm{d}\boldsymbol{k}_4 \left\langle \tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_1)\tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_2)\tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_3)\tilde{\delta}^{(1)}_{\rm mass}(\boldsymbol{k}_4)\right\rangle G(\boldsymbol{k}_3,\boldsymbol{k}_4,z'',z''')e^{\mathrm{i}\boldsymbol{k}_3\cdot\boldsymbol{r}''(z'')}e^{\mathrm{i}\boldsymbol{k}_4\cdot\boldsymbol{r}''(z''')}\\ \times e^{\mathrm{i}\boldsymbol{k}_1\cdot\boldsymbol{r}(z)}e^{\mathrm{i}\boldsymbol{k}_2\cdot\boldsymbol{r}'(z')} = (2\pi)^2\int \mathrm{d}^2\boldsymbol{k}_1\mathrm{d}^2\boldsymbol{k}_2 P^l_\delta(k_1)P^l_\delta(k_2)G(\boldsymbol{k}_1,\boldsymbol{k}_2,z,z')W(|\boldsymbol{k}_1\mathcal{D}_0(z'')+\boldsymbol{k}_2\mathcal{D}_0(z''')|\theta)\\ \times W(k_1\mathcal{D}_0(z)\theta)W(k_2\mathcal{D}_0(z')\theta)\delta_D(\chi(z'')-\chi(z))\delta_D(\chi(z''')-\chi(z'))+\mathrm{perm}(z\leftrightarrow z').\end{aligned} \tag{1.78}$$

Making the change of variable $\boldsymbol{k}_1\mathcal{D}_0(z'')\rightarrow\boldsymbol{\ell}_1$ and $\boldsymbol{k}_2\mathcal{D}_0(z''')\rightarrow\boldsymbol{\ell}_2$ allows us to express the correction to the third order moment as

$$\begin{aligned}\left\langle \left(\kappa^{(1)}\right)^2\kappa^{(2)}_{\rm corr}\right\rangle_\theta = -\frac{4}{(2\pi)^4}\int_0^z\frac{\mathrm{d}z'}{H(z')}\omega(z',z)^2\int_0^{z'}\frac{\mathrm{d}z''}{H(z'')}\omega(z'',z)\omega(z'',z')D_+(z')^2D_+(z'')^2\\ \int\frac{\mathrm{d}\boldsymbol{\ell}_1\mathrm{d}\boldsymbol{\ell}_2}{[\mathcal{D}_0(z')\mathcal{D}_0(z'')]^2}P^l_\delta\left(\frac{\ell_1}{\mathcal{D}_0(z')}\right)P^l_\delta\left(\frac{\ell_2}{\mathcal{D}_0(z'')}\right)H(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2)W(\ell_1\theta)W(\ell_2\theta)W(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|\theta),\end{aligned} \tag{1.79}$$

where

$$H(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2)=\frac{\boldsymbol{\ell}_1\cdot\boldsymbol{\ell}_2}{\ell_2^2}+\frac{(\boldsymbol{\ell}_1\cdot\boldsymbol{\ell}_2)^2}{\ell_1^2\ell_2^2}. \tag{1.80}$$

Finally, using the following properties (Bernardeau, 1995) of the 2D top-hat window function where φ is the angle between the two wave-modes $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2$

$$\int_0^{2\pi}\mathrm{d}\varphi W\left(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|\right)\left[1-\cos^2(\varphi)\right]=\pi W\left(\ell_1\right)W\left(\ell_2\right), \tag{1.81}$$

$$\int_0^{2\pi}\mathrm{d}\varphi W\left(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|\right)\left[1+\cos(\varphi)\frac{\ell_1}{\ell_2}\right]=2\pi W\left(\ell_2\right)\left[W\left(\ell_1\right)+\frac{\ell_1}{2}W'\left(\ell_1\right)\right], \tag{1.82}$$

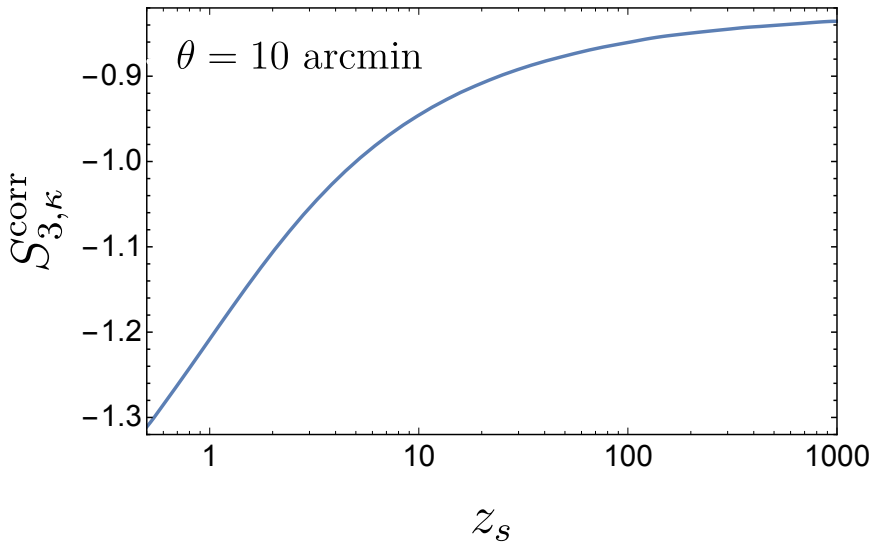


Figure 1.3: Leading order correction to the skewness induced by lens-lens couplings and geodesic deviation. The field is filtered in top-hat window of angular radius $\theta = 10$ arcmin and we plot the evolution of the correction with respect to the source redshift. This figure was published in **Barthelemy** et al. (2020b).

and plugging them in equation (1.79) finally yields

$$\left\langle \kappa^3_{\rm corr} \right\rangle_\theta = -\frac{48}{(2\pi)^2}\int_0^{z_s}\frac{{\rm d}z'}{H(z')}\omega(z',z_s)^2\int_0^{z'}\frac{{\rm d}z''}{H(z'')}\omega(z'',z_s)\omega(z'',z')D_+(z')^2D_+(z'')^2$$
$$\int\frac{{\rm d}\ell_1{\rm d}\ell_2}{\left[\mathcal{D}_0(z')\mathcal{D}_0(z'')\right]^2}P^l_\delta\left(\frac{\ell_1}{\mathcal{D}_0(z')}\right)P^l_\delta\left(\frac{\ell_2}{\mathcal{D}_0(z'')}\right)\frac{J_1(\ell_1\theta)(J_0(\ell_1\theta)-J_2(\ell_1\theta))J_1(\ell_2\theta)^2}{\ell_2\theta^3}, \quad (1.83)$$

where the subscript $_s$ now denotes the source plane. This notation is introduced since no distribution of sources are here taken into account for implementation simplicity and notation clarity but, as already shown previously in section 1.5.4, the generalisation is nevertheless straightforward. The integral in equation (1.83) can then be computed numerically.

1.7.3 Plots and evaluation of the correction

For an opening angle $\theta = 10$ arcmin, that is the angular radius of the top-hat window, I obtain the correction on the skewness due to lens-lens coupling and geodesic deviation shown in figure 1.3 as a function of the source redshift. As implied in the derivation, this is not the complete correction including all sources of couplings between lenses since I only considered the first non-zero (leading) term beyond the linear regime. As expected, the correction for this scale and up to high source redshifts is shown to be very small compared to the standard perturbation theory (SPT) tree-order order skewness* displayed in figure 1.4 (blue solid line). It barely reaches a few percent of the signal at low redshifts. However, the correction becomes relatively more important as the source redshift increases since it appears to be relatively

*I intentionally did not dedicate a section to the derivation of this quantity, since it will be done in the following chapters making use of the projection formula (3.4) and large deviation theory for the matter density field. It is nevertheless a very close computation to the one that I presented here replacing the kernel G (1.73) by that accounting for second order perturbation theory (Bernardeau et al., 2002).

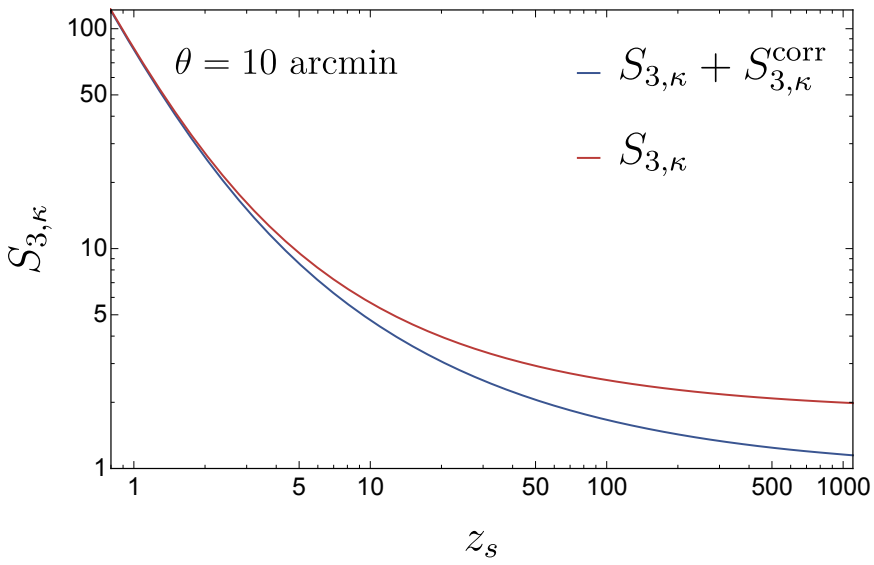


Figure 1.4: Leading order skewness with (blue) and without (red) considering the correction induced by lens-lens couplings and geodesic deviation. The field is filtered with a top-hat window of angular radius $\theta = 10$ arcmin and we plot the evolution of the skewness with respect to the source redshift. This figure was published in **Barthelemy** et al. (2020b).

constant in time/redshift, whereas the tree-order skewness value roughly decreases as $z_s^{-1.35}$ (Bernardeau et al., 1997). The effect is thus proven to be more important at high redshift, especially in the context of CMB lensing for which the post-Born corrections have the same amplitude as the signal itself and tend to make the field even more Gaussian (with a skewness going roughly from 2 to 1 for the aperture of 10 arcmin used here).

I also show in figure 1.5, in the case of a source at $z_s = 1100$ and an opening angle of $\theta = 10$ arcmin, the contribution to the skewness due to the couplings between the lenses at redshift z and z'. This contribution is null when the lenses are at the same redshift or one of them is at the observer or the source i.e $z' = 0$ or $z = z_s$ and peaks at relatively low redshift $(z, z') \sim (1.5, 0.7)$.

Finally, figure 1.6 compares measurements of the skewness in the simulated convergence maps of Takahashi et al. (2017) that I present in detail in the next section 1.8 (blue points with error bars given by the dispersion among the mock data) to the theoretical prediction from tree-order standard perturbation theory with (blue solid line) and without (red solid line) post-Born corrections. For sufficiently large opening angles ($\theta \gtrsim 10$ arcmin), the post-Born corrected prediction is in full agreement with the simulation. At smaller scales, deviations start to appear and the measured value of the skewness overshoots the prediction, an effect which might be due to non-linearities not taken into account in the tree-order result and probably calls for additional (higher order) corrections both in the dynamics of the matter field and post-Born effects. Note however, that the high-order cumulants themselves are not very well tested in the simulation and that the simulated cosmic variance is relatively large as can be seen from the size of the error bars, and we are thus prevented from testing the prediction very precisely. Additionally, when comparing theory and simulated data, let us keep in mind that the error bars for different angular scales are highly correlated, so

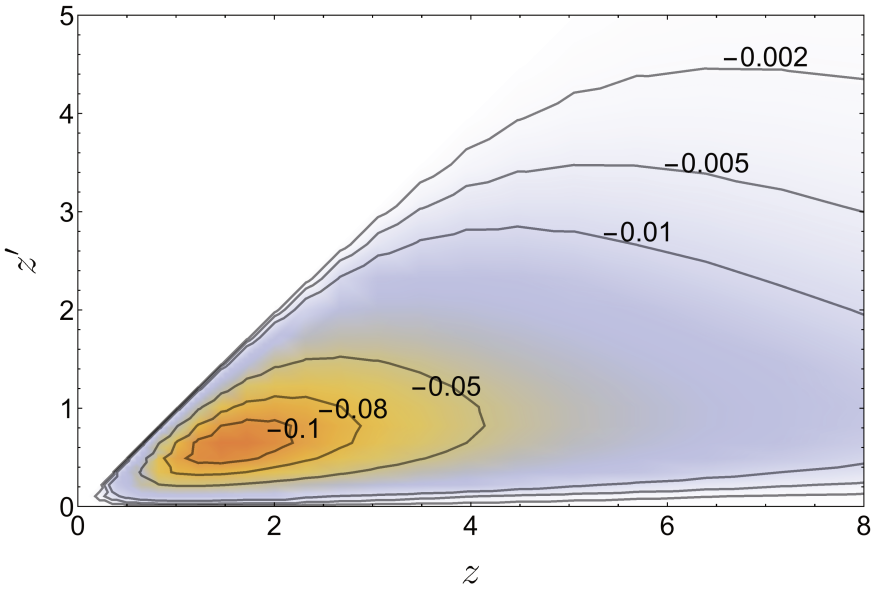


Figure 1.5: Contribution to the correction of the skewness from couplings between lenses (including the so-called lens-lens coupling and geodesic deviation) at redshift z and z' for the CMB convergence field and an aperture $\theta = 10$ arcmin. This figure was published in **Barthelemy** et al. (2020b).

that overall there is no significant deviation of the measurements from the prediction. It nonetheless would be very interesting to test in more detail the numerical simulation used here although this is not the purpose of this section.

1.8. Numerical simulations of weak lensing fields

Numerical simulations are not central in the work presented in this thesis, but they are still rather important, both as a method prolifically used in the community and as a means for researchers like myself to provide intuition, to provide quantities that we are not (hopefully yet) able to analytically compute, and to make comparisons that both help us quantify the range of validity of our computations and conversely test details of the simulations that would otherwise be difficult to test in the absence of theoretical models. I aim in this section to give a brief description of the methods that are used to simulate weak-lensing maps in the context of cosmological computations and hopefully provide insight on their interconnection with our theoretical models.

1.8.1 Motivating numerical simulations

The lensing formalism that I introduced in this chapter until now relied on pure theoretical approaches where I either justified or assumed that the approximations and perturbation strategy employed were sufficiently correct or accurate. However, there are a number of issues that limit the sole usage of theoretical approaches for real surveys and where the use of numerical solvers for the dynamics at play in the universe seems more appropriate.

First of all, as is apparent in figure 1.7 for general lensing quantities or equation (1.38) for the convergence power spectrum, an accurate knowledge of the clustering of the late-

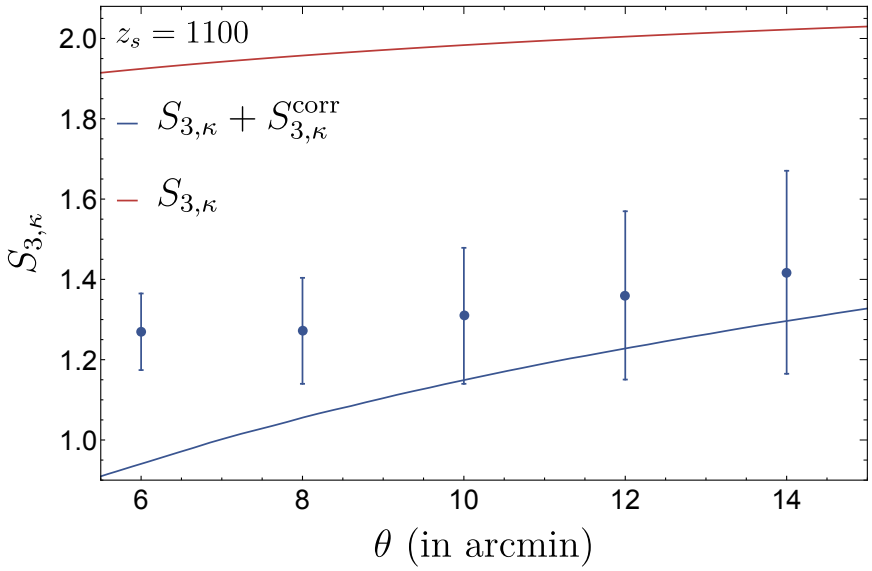


Figure 1.6: Convergence skewness of the CMB with respect to the opening angle of the top-hat smoothing applied. In red (resp. blue) is the skewness evolution as computed from large deviation theory without (resp. with) taking into account post-Born corrections. The measured points are from 23 simulated maps (mean of the 23 realisations) described in the main text and error bars represent their standard deviation around the mean. This figure was published in **Barthelemy** et al. (2020b).

time small scales (*i.e.* very non linear scales) of the density field is required even for the description of intermediate angular scales ($\sim 2 - 10$ arcmin). However, this proves to be a challenge for most* theoretical approaches, for the power spectrum of course but even more so for high-order statistics like the bi-spectrum, peak counts or the PDF. On the other hand, dark-matter N-body simulations allow for very accurate, high-resolution simulations of the matter density field that span a wide range of scales.

Numerical simulations are also a great tool with which to potentially generate hundreds or even thousands of independent realisations (one example of a recent massive N-body simulation is Villaescusa-Navarro et al., 2020; Petri et al., 2016a, is rather a paper specific to lensing that tries to estimate how independent lensing maps generated from a same smaller N-body box are) of the same observable, which can then be used to estimate its covariance matrix, which for example includes its cosmic variance. This estimation, often hard to perform from a theoretical point of view, is nonetheless key to the assessment of errors in cosmological analysis and will be better described and used in section 3.8.

For theorists, simulations are an interesting way to test the validity range of some of the mathematical approximations used throughout this chapter such as the Born-and-independent-lenses approximation, the reduced shear approximation, and even the perturbation order at which the density field is computed. However, this affirmation is not valid for every observable in every regime and will be nuanced in section 1.8.3.

Another very important (and already discussed in section 1.5.4) point, especially in the context of cosmological weak-lensing, is the inclusion of baryonic effects – processes that

*I present in section 1.9 one theoretical method, the BNT transform, that precisely aims at separating physical scales in weak-lensing observables and which is thus nonsensitive to this issue.

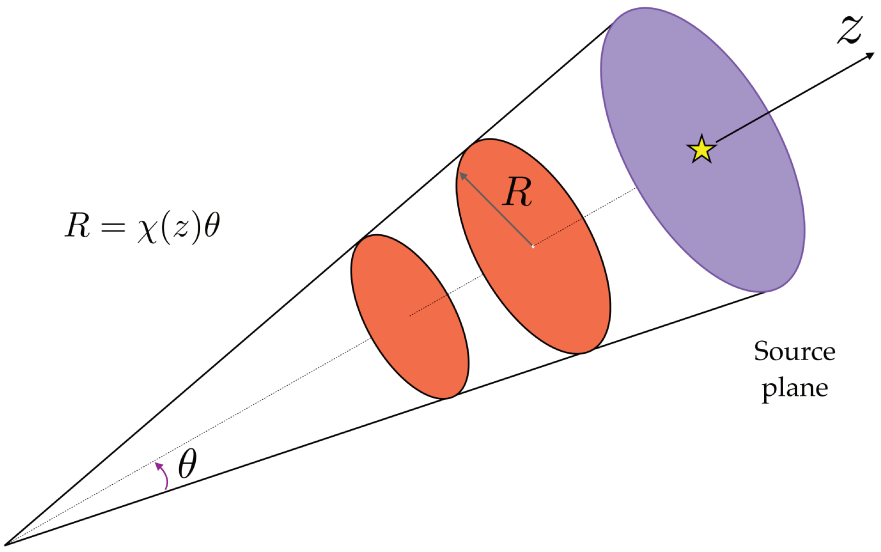


Figure 1.7: Schematic description of the line of sight integral leading to the filtered convergence.

heat and cool gas, re-distribute it or transform it into stars – in our models. Indeed, those have an impact at small and intermediate scales on the clustering of matter (see for example figure 2 of Schneider et al. (2019a)), which thus affect lensing, while other physical effects treated as systematics in lensing experiments such as the intrinsic alignment of galaxies and the clustering between sources and lenses are also affected, or even a direct consequence because there would not be galaxies without baryonic physics. The inclusion of those effects is performed in hydrodynamic simulations (two very well known example include Teyssier, 2002; Springel, 2010), where in addition to dark matter particles, the gas is seen as an ideal fluid obeying Euler equations and additional physics necessary for realistic simulations but happening at scales smaller than those resolved are also implemented. I have already briefly described the challenges with these "sub-grid" effects in section 1.5.4, but let me repeat that they follow semi-analytical phenomenological models that include radiative cooling, star formation, supernova feedback, magnetic fields, black hole and AGN feedback, and cosmic rays, and are often not well-known in detail, which significantly complicates the precise evaluation of their impact on lensing observables (Naab & Ostriker, 2017). Moreover, these hydro-simulations are very computationally expensive, and hence simulations that resolve galaxies are usually performed in volumes too small (~ 100 Mpc$/h$ and thus in the absence of large modes and or with inadequate statistics) for cosmological application. One idea is then to parametrise and fit mass profiles in these hydro-simulations (or directly if, for example in X-ray, measurements are available) and use these to displace dark matter particles in pure N-body simulations as is done for example in Schneider et al. (2019a). Other approaches include halo-model-infused fitting functions for the power spectrum, but these still require input from numerical simulations as in Mead et al. (2020).

Finally, numerical simulations are a very appropriate method to take into account a multiplicity of specific survey constraints that are otherwise hard to take into account. Those constraints include for example inhomogeneous depth in the field of view or a complicated mask geometry. Simulations can also account for complex interplay between astrophysics

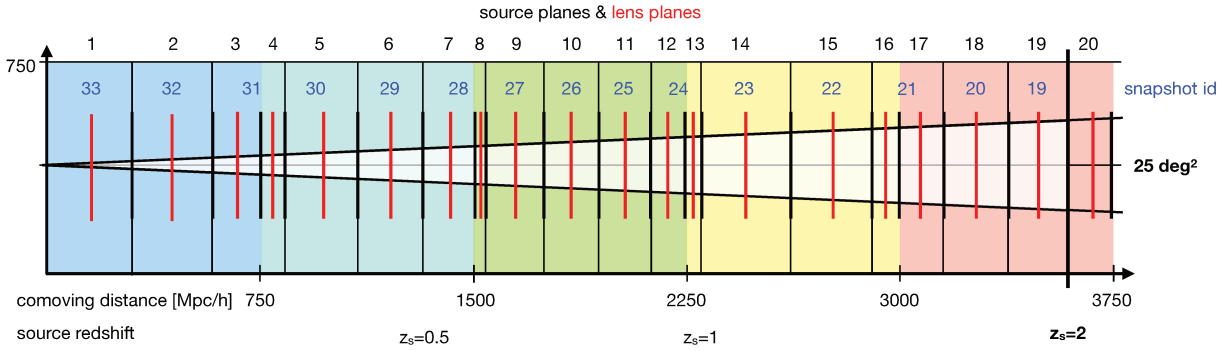


Figure 1.8: Illustration explaining how the DUSTGRAIN light-cones up to source redshift $z_s = 2$ are built from a single realisation N-body simulation of box size 750 Mpc/h. To reach the source redshift of $z_s = 2$, the box is replicated 5 times along the light-cone up to $z_s = 2$. A randomisation procedure is applied to avoid replicating the same structure along the line of sight (same colour indicates same randomisation) and generate 256 realisations. This figure was published in Boyle et al. (2021).

and cosmology, like when in high-density regions, galaxy pairs that are formed are ignored because their images are blended and thus ignored by standard lensing pipelines which leads to an under-representation of those high-density regions (Hartlap et al., 2011). A similar issue is present when massive foreground galaxies block the line of sight to background galaxies.

1.8.2 Implementation: N-body simulations and ray-tracing

Though there exist some counter-examples that I will briefly mention, and though codes that implement the same idea may differ in accuracy and performance (Hilbert et al., 2020), a particular scheme to produce cosmological simulated weak-lensing maps stands out. I am focusing here on pure dark-matter N-body simulations, but the same ideas, at least for the ray-tracing, can also be applied to hydro-simulations. The basic approach is to produce a (set of) box(es) of purely dark matter particles, evolve it (them) following Newtonian dynamics between the particles and store snapshots at different redshifts. From there, the light-cone is reconstructed from the simulated snapshots from the observer to the sources, and light rays are re-traced from the observer to the sources by considering discrete thick lens planes/shells (particles are smoothed and projected onto density planes) where light moves in straight paths and is only deflected when encountering a new lens plane/shell. This multiple-lens-plane approach thus takes into account couplings between the lenses and generates a non-symmetric Jacobian matrix A. The ray tracing allows reconstruction of the projected gravitational potential and thus the convergence, shear and rotation. In the following, I will describe two specific simulation suites that I use in this thesis, the DUSTGRAIN-pathfinder simulation that is used when investigating the cosmological information content of the convergence PDF thanks to its maps with varying cosmologies, and the Takahashi full-sky lensing simulations, which I use as a comparison point for the theoretical formalism that I present for the convergence and aperture mass PDFs.

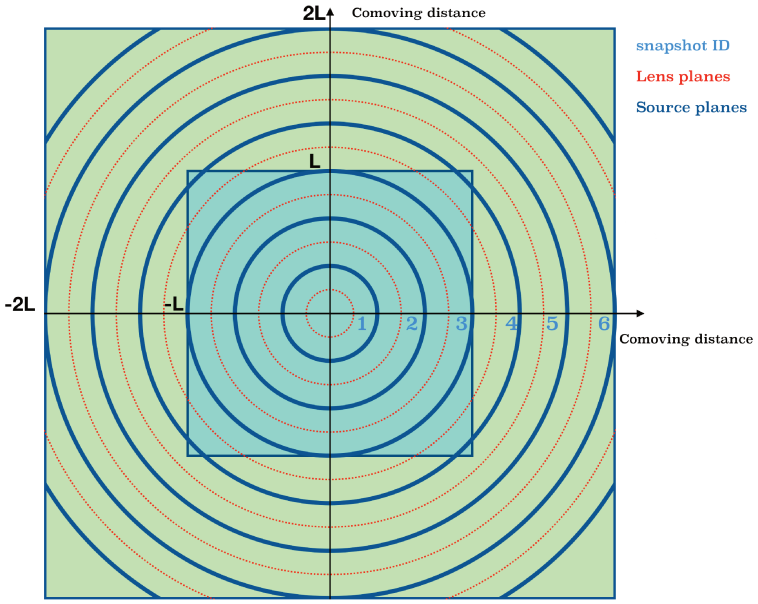


Figure 1.9: Illustration explaining how the Takahashi light-cones are built from a series of 14 nested N-body simulations of box size multiple of $L = 450$ Mpc/h. There are 6 independent copies of each box and for each 6 series of 14 boxes, the centre of each box is redefined 18 times leading to $18 \times 6 = 108$ realisations in total. The same colour in each box indicates the same randomisation.

DUSTGRAIN: The DUSTGRAIN-pathfinder simulations (Giocoli et al., 2018) are a suite of cosmological simulations performed with the MG-GADGET N-body code and subsequently post-processed with the weak-lensing map making code MapSim (Giocoli et al., 2014). It notably includes a set of standard ΛCDM cosmologies with specific deviations of individual cosmological parameters from their fiducial values. The simulations follow the evolution of 768^3 dark matter particles in a periodic cosmological volume of 750 Mpc/h per side, starting from initial conditions generated at $z = 99$ from a random realisation of an initial matter power spectrum computed through the CAMB Boltzmann solver (Lewis & Bridle, 2002) within the Zel'dovich approximation (that is first order in lagragian perturbation theory or 1LPT). The light-cone is built using the geometry illustrated in figure 1.8. From a single realisation of the N-body simulation there are 21 snapshots available. Given the box length of 750 Mpc/h, roughly 5 boxes are needed to cover the comoving distance of about 3.6 Gpc/h to $z_s = 2$. To obtain better redshift sampling, each simulation volume necessary for constructing the light-cone is divided along the line of sight into multiple redshift slices obtained from the individual snapshots. If the redshift slice reaches beyond the boundary of a single box, two lens planes are constructed from a single snapshot. The total number of lens planes up to $z_s = 2$ is 19. To avoid replicating the same structure along the line of sight, the 5 boxes needed to cover the light-cone are randomised which is possible thanks to the periodic boundary conditions of the box. This randomisation procedure allows to extract multiple realisations from a single simulation and is achieved by using seeds that act on the simulation boxes based on

- changing the location of the observer, typically placed at the centre of one of the faces of the box,

- redefining the centre of the box (taking advantage of periodic boundary conditions), and
- changing the signs of the box axes.

Note that the same randomisation process is applied to all redshift slices belonging to the same box to avoid spatial discontinuities. The result is a set of flat cartesian maps with square pixels which makes sense only thanks to the fact that the patches are small enough that the curvature of the sky can be ignored.

TAKAHASHI FULL-SKY LENSING SIMULATIONS: The Takahashi et al. (2017) simulation suite consists of 108 state-of-the-art full-sky gravitational lensing simulation data sets. Note that the simulations being full-sky (or at least 1/8 of the sky because of box replication) is an added bonus in the comparisons to our theoretical formalism since it equates to sufficient statistics given by a large simulated volume and the presence and impact of long(er) wave-modes which are absent when considering only small patches. Fourteen boxes of pure dark matter particles with side lengths of $L = 450\text{Mpc}/h$, $2L \cdots 14L$ were prepared along with 6 independent copies. The number of particles for each box was 2048^3, making the mass and spatial resolutions better for smaller boxes but the overall angular resolution (number of particles per surface area) roughly constant at each redshift. Each box was evolved in a periodic cosmological N-body simulation following the gravitational evolution of dark matter particles without baryonic processes using the N-body code GADGET2. The initial conditions were generated from second-order Lagrangian perturbation theory (2LPT) with the initial linear power spectrum calculated once again using CAMB. Ray-tracing was performed using the public code GRAYTRIX which follows the standard multiple-lens plane algorithm in spherical coordinates using the HEALPIX algorithm. The light-cone is build as illustrated in figure 1.9. The boxes were placed around a fixed vertex representing the observer's position while each box was duplicated eight times and placed around the observer using periodic boundary conditions. Spherical lens shells with width of 150 Mpc/h (3 per box) were then considered to trace the resulting light-ray paths from the observer to the source planes. To increase the number of realisations of source planes at each redshift, the centre of each box was redefined 18 times, again this is possible thanks to the periodic boundary conditions in the boxes, resulting in $18 \times 6 = 108$ realisations. The data set eventually includes full-sky convergence maps from redshifts $z = 0.05$ to 5.3 at intervals of 150 Mpc/h comoving radial distance and are freely available for download*. The pixelization of the full-sky maps follows the HEALPIX ring scheme with available resolutions of NSIDE = 4096, 8192 and 16384, a specific method to pixelise the sphere with pixels of same area $2^{nside}/4\pi$. Note that the data set also includes lensing map at the last scattering surface ($z = 1100$ which is also the redshift of the CMB). Those were obtained making full use of the N-body simulation up to redshift ~ 7 and generating a density field of the earlier epochs based on the linear power spectrum calculated using CAMB.

The ray-tracing techniques presented in the two simulations are very similar and particularly used in modern simulations. Still, it might be useful to mention that some older

*http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/

simulations used to work within the Born approximation, just projecting the density field onto source planes to obtain the convergence field and then subsequently the projected gravitational potential and the shear. Another method introduced in Gouin et al. (2019) consists of obtaining directly the deflection angle of a light ray by integrating along the line of sight the transverse acceleration (thus the transverse derivative of the gravitational potential) of encountered cells. This offers the advantage that the very same gravitational field that was used to move particles (and in their case evolve Eulerian quantities) is therefore used to consistently derive the deflection field. In particular, the way shot noise is smoothed out in the simulation to recover the acceleration field from a mixture of Lagrangian particles and Eulerian gas cells (this is the case in the Ramses simulation that they use) is faithfully respected in the ray-tracing, which means that the force felt by photons is very similar to the one felt by particles in the simulation. These last few years have also seen a gain in interest in directly solving the general relativistic geodesics equations for the light path from the observer to the sources. This has made possible interpreting the positions of the Newtonian N-body dark matter particles as being expressed in the N-body gauge (Fidler et al., 2016) and holds the advantage of unifying the deflection, lensing and redshift calculations. This is notably very interesting for studying redshift space distortions since every contribution to the "measured redshift", comoving redshift (what we usually mean when mentioning redshift), Doppler effect coming from peculiar velocities, gravitational redshift gained from differences in the gravitational potential along the light path, and so on, is then accounted for and can be compared to the "real" comoving redshift that we use both as a measure of time and distance in cosmology and that is also given in the simulation (Breton et al., 2019). Otherwise, the redshift space distortions are often included in lensing numerical simulation by considering perturbations in the source planes only taking into account peculiar velocities.

1.8.3 Known limits in our context

As we have seen, there are many good reasons why numerical simulations are an essential tool to better understand and model different aspects of cosmic shear experiments. However, their apparent ability to access any quantity or considerably reduce error bars providing they have enough volume and resolution could be a bit misleading. In the context most useful to this work, that is to validate products of a theoretical formalism without baryonic physics, some limitations appear at the level of observables beyond the standard 2-point statistics (correlation function and power spectrum), which are almost always measured and compared to previous theoretical formalism or ansatz. Indeed many recent works started the daunting task of evaluating the accuracy of weak-lensing simulations for those non-Gaussian statistics, starting by comparing different codes between themselves (Hilbert et al., 2020), and it for example turns out that some fine-tuning of the algorithms are then needed for ray-tracing codes to reproduce the same convergence PDF with percent accuracy from the same light-cone. Matilla et al. (2020) also studied the impact of the thickness of the lens planes used to build past light-cones and the mass resolution of the underlying N-body simulation and thus proposed guidelines for the design of future numerical suites but which are not implemented in the – still state-of-the-art – numerical simulations that I am using in this thesis. Another issue is the pseudo-independence of lensing maps generated from randomisation of the line of sight through replication of a small but better resolved N-body box. As stated before, Petri

et al. (2016a) for example suggested that maps generated from 1 or 2 realisations of the same box could be considered independent but those results, to my knowledge, were not replicated nor extended to every set of observables such as the convergence PDF or moments and even less for the aperture mass or the shear. Numerical uncertainties of the density field itself might also have a strong impact on the prediction of non-Gaussian aperture mass statistics as I later show in the case of the $M_{\rm ap}$ skewness in the Takahashi simulation in section 4.3. This might be explained by the fact that the aperture mass is seen as the difference between the convergence field at two different scales (see equation (4.1)), and thus the difference between numbers of similar amplitude, which relatively gives an important weight to the numerical uncertainties associated with those two values. Building light-cones from the association of different N-body boxes also assumes that frequent discontinuities of structures have a negligible impact on observables. One could assume that this might indeed be the case when considering very large lensing kernels over a wide variety of structures but this impact is not to my knowledge quantified and might be bigger when narrowing the lensing kernel as is the case when applying the BNT transform that I present in the next section 1.9. Finally, another important issue that must be taken into account, although not present for the Takahashi simulation, is the impact that small maps/patches constructed as planar projections of part of the sphere have on high-order statistics of the lensing fields. Such a study was recently performed in Vallis et al. (2018) and indeed showed a measurable impact on peak counts and Minkowski functionals. All of this suggests that indeed more work is needed if we are to extract and trust subtle non-linear effects from weak-lensing simulated maps.

On the other hand, this also gives a new light to the development of so-called "from first principles" theoretical approaches in the sense that they provide means to test, at least in some regime, high-order/non-Gaussian statistics in numerical simulations. This creates an osmosis between the two approaches which are thus seen as co-dependent to gain insight on the large-scale structure of our Universe.

1.8.4 Post-processing of convergence maps

To conclude this section on numerical simulations, I will just briefly explain how I post-process the simulated convergence maps that I obtain from the simulations. I use exactly the same methods (in real and harmonic space) for both DUSTGRAIN and Takahashi's simulations but I only describe it for healpix (Takahashi's) maps, which are less frequent. I also illustrate the smoothing procedure by showing a small patch of a full-sky convergence map from the Takahashi simulation in figure 1.10.

HEALPIX MAPS: To convolve those maps with top-hat windows of the desired angular radii and thus access simulated κ and $M_{\rm ap}$ statistics, I use two different methods that proved to give equivalent results at more than the percent precision for the variance, skewness and kurtosis, and indistinguishable by eye for PDFs. My *real-space* method consists in using the *query_disc* function of HEALPY, a python package specialised in the processing of healpix maps, to find all pixels whose centres are located within a disk centred at one specific pixel p, which then allows to reassign the value of p as being the mean of all the pixels inside the disk. My *dual-space* method consists in convolving the map with the appropriate filter

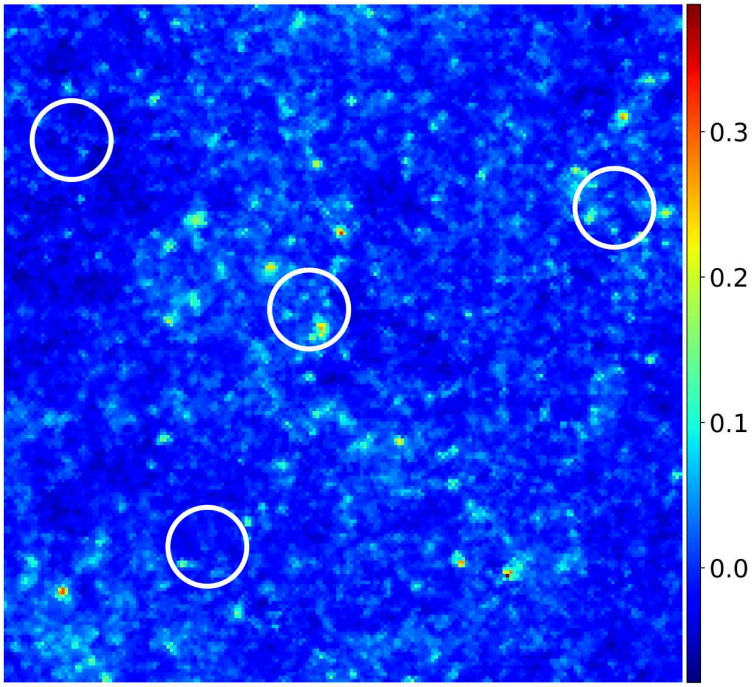


Figure 1.10: Gnomonic projection of a 172×172 arcmin^2 piece of the convergence field at redshift 2 taken from 1 realisation of the Takahashi simulation. The resolution per pixel is 0.86 arcmin. The white circles show an example of a top-hat smoothing of angular radius of 10 arcmin. This figure was published in **Barthelemy** et al. (2020a).

by decomposing both the map and filter in spherical harmonics using the *map2alm* and the *beam2bl* functions of HEALPY, convolving the map and filter in this space (a simple multiplication) and then going back to pixel-space via the *alm2map* function. The two methods were found to agree at the level of each map and for all scales which was not a given since i) the *real-space* procedure does not yield an exact top-hat and ii) the a_{lm} transform has a non empty kernel, meaning that a randomly generated map in pixel-space, sent to a_{lm}-space and back might be significantly different from the original one.

1.9. The Bernardeau-Nishimichi-Taruya (BNT) transform: geometrical sorting of scales and dynamics along the line of sight

Let me now suppose that I want to analyse lensing data coming from a modern survey such as Euclid or LSST. Regardless of the model for the statistics I choose to consider, I know for a fact that this model – either theoretical, extracted from numerical simulations or a mix of the two – will have some limitations, the most usual being its validity range for physical scales and redshifts on the dynamics of matter, the confidence in the baryonic model considered, its compatibility with the introduction of other systematics, etc. Hence to extract information from the observed maps, some sort of cuts might be considered in order to make sure that we are not over-interpreting data while still keeping the maximum amount of detail on what was observed. In the case of lensing, these sorts of thoughts are made complicated by the fact that the line of sight projection leading to smoothed lensing fields (past light-cone integration),

and exemplified in figure 1.7, probes and mixes different scales – small at the tip and large close to the source – and structures at different times/redshifts.

As such, the cuts I would have to perform would be sub-optimal in the sense that to make sure that my model makes sense I am thus obliged to consider large enough angular scales and/or higher source redshifts so as to dilute the influence of the smallest physical scales and redshifts that contribute to the lensing effect at this angular scale. This in particular renders the use of theoretical descriptions of lensing quantities much more difficult since the limitation on accessible physical scales is even more apparent for standard techniques, *i.e.* perturbation theory. Weak-lensing statistical probes are thus often modelled by more phenomenological approaches such as halo models that can also take into account baryonic physics which becomes important at small scales (Mead et al., 2020), and even more so making use of numerical simulations (Schneider et al., 2019a, and see sections 1.5.4 and 1.8). However and as already mentioned in this chapter, and more specifically the aforementioned sections, those simulations are not always tested in fine detail for high-order non-Gaussian statistics and still suffer limitations at small scales, for example coming from the influence of baryonic physics.

Alternatively, a theoretical strategy to disentangle physical scales in lensing quantities known as the Bernardeau-Nishimichi-Taruya (BNT) transform or nulling strategy was proposed by Bernardeau et al. (2014b) and notably re-enables the use of pure theoretical tools for accurate predictions. Note that this idea of nulling inexploitable scales where theoretical models cannot be trusted is not particularly new and can to my knowledge be traced back all the way back to Huterer & White (2005) on the power spectrum and Joachimi & Schneider (2008) on the specific effect of intrinsic alignments. However, the focus here could rather be phrased in terms of re-organising the information coming from different physical scales rather than getting rid of it as will be apparent in the next subsections. Nothing is lost or technically nulled. Recent examples of the application of this strategy in the literature can for example found in Taylor et al. (2018) and Taylor et al. (2021) to remove the sensitivity to the poorly modelled small scales for the two-point cosmic shear signal, and therefore improve cosmological constraints using the Dark Energy Survey shear data.

1.9.1 Intuitive idea

We have already noticed that within the Born approximation with independent lenses the convergence (1.58) is expressed as a weighted integral along the line of sight of the underlying matter density contrast, that is the (non-linear) perturbations occurring on the background. However the weighting function (lensing kernel) is expressed purely in terms of background geometrical quantities: the fraction of matter in the energy budget of the universe Ω_m, the angular comoving distance $\mathcal{D}_0$ and the redshift z. As such, attributing a zero weight to the lenses close to the observer and that contribute to the very non-linear part of the signal is a purely geometrical problem independent of our ability to describe the non-linear dynamics taking place at those scales and redshifts.

The idea of the BNT transform is thus at the same time brilliant and quite simple. Say we were to observe the convergence or the shear fields* of sources located at different redshifts.

*The formalism is here described for the convergence but is valid for the shear given the relations between the two.

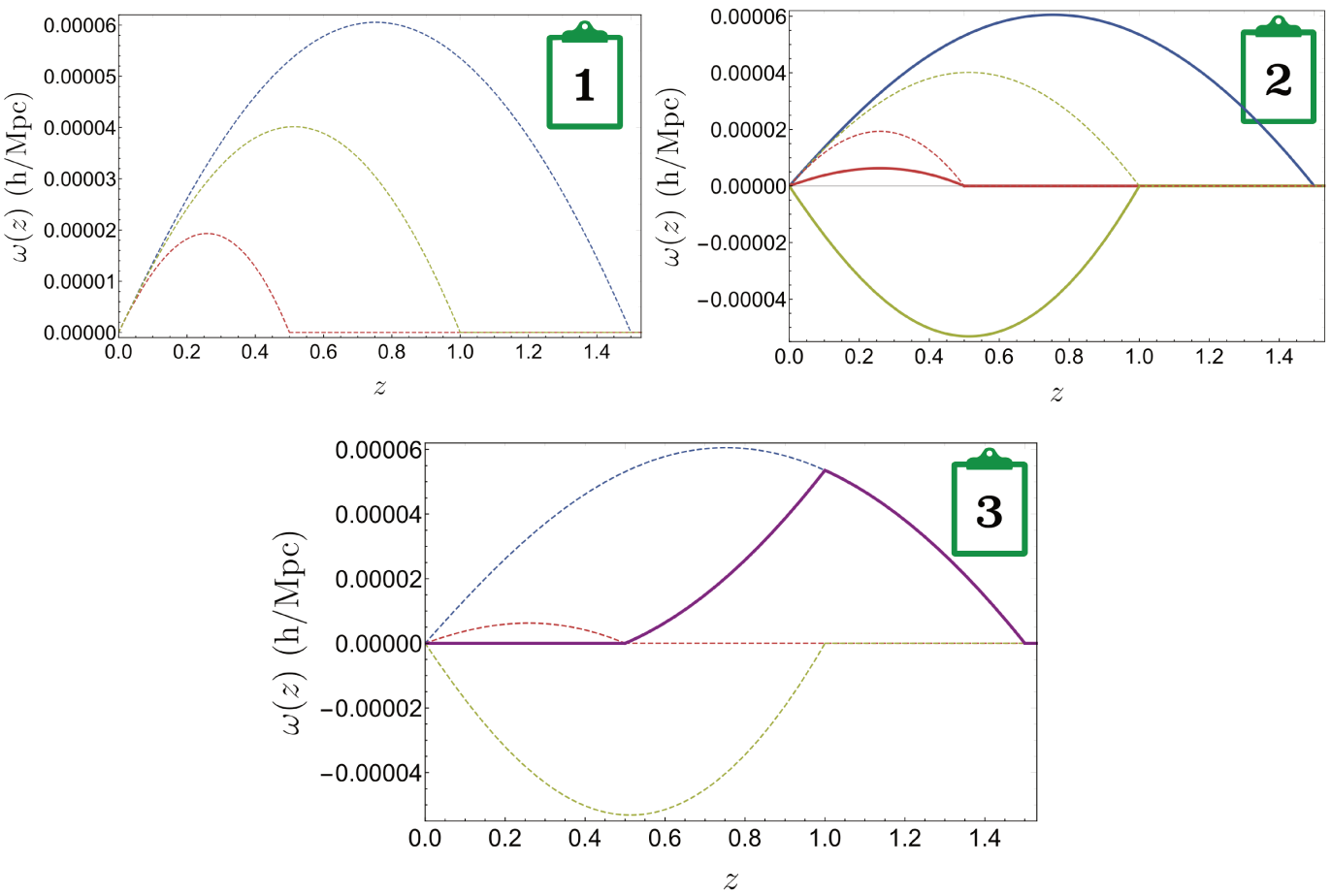


Figure 1.11: Illustration of the BNT transform. The first panel displays 3 lensing kernels for sources located at 3 different redshifts. The 2nd one shows how those kernels evolve once we multiply them by carefully chosen real numbers, the BNT weights, and the weight associated to blue kernel is 1. The 3rd panel finally shows the resulting sum of the 3 modified kernels as the thick purple line. One can thus see how this operation effectively nulled the influence of all lenses between the observer and the closest sources.

Their linear combination would obviously still be an observable since it would amount to simply summing observed maps with chosen weights. However, at the level of the theoretical description, the resulting lensing field would still be described by a weighted integral along the line of sight of the matter density contrast, where the resulting *BNT lensing kernel* is now simply expressed as the linear combination of the initial lensing kernels. Now, should we be able to find scale and redshift-independent weights such that the resulting BNT kernel is null at the scales inaccessible to a good theoretical description, say between the observer and the first observed sources, then we would have effectively nulled their influence, recovering our ability to provide a theoretical description of lensing quantities. I illustrate this idea in figure 1.11 and another illustration of how this extends our ability to describe small angular scales on the shear power spectrum is presented in Bernardeau et al. (2014b).

1.9.2 Source plane formalism

Now that the intuitive description of the previous subsection hopefully managed to make things clear, let me continue by describing in more detail the formalism behind this transformation.

As mentioned earlier, the principle of this nulling transformation is to linearly combine successive convergence maps so as to define a new observable to avoid the contribution of many physical scales for one fixed angular scale. Starting from several κ-maps at redshifts z_i, the BNT or nulled convergence is defined by

$$\kappa(\boldsymbol{\theta})_{\rm null} = \sum_i p_i \kappa_i, \tag{1.84}$$

where p_i are dimensionless weights whose values are chosen so as to reach the desired effect. The resulting BNT lensing kernel is thus written as

$$\omega(z)_{\rm null} = \frac{3\Omega_m H_0^2}{2c^2} \sum_{i,z_i>z} p_i \frac{\chi(z)\left(\chi(z_i)-\chi(z)\right)}{\chi(z_i)}(1+z), \tag{1.85}$$

where $\chi(z) = \mathcal{D}_0(z)$ in flat FLRW spacetime and is the comoving radial distance. The game now amounts to finding a set of p_i so that the built BNT convergence map is only sensitive to lenses confined in a certain range of distances for which an exquisite knowledge of the small-scale physics is not necessary and standard theoretical predictions, such as those of SPT, are accurate.

Fortunately, such solutions for a set of discrete planes exist. They are unique for sets of 3 source planes up to a normalisation constant and the generalisation to sets of more than 3 source planes are obtained by considering sets of 3 planes at a time. Let me for instance assume that three source planes are located redshifts z_i, $i = 1, 2, 3$ and notice that the expression of $\omega(z)_{\rm null}$ can be re-written as

$$\omega(z)_{\rm null} = \frac{3\Omega_m H_0^2}{2c^2}(1+z)\chi(z)^2 \left[\frac{1}{\chi(z)} \sum_{i,z_i>z} p_i - \sum_{i,z_i>z} \frac{p_i}{\chi(z_i)}\right]. \tag{1.86}$$

Now it appears more clearly that if the weights p_i satisfy the two conditions

$$\begin{cases} \displaystyle\sum_{i=1}^{3} p_i = 0, \\ \displaystyle\sum_{i=1}^{3} \frac{p_i}{\chi(z_i)} = 0, \end{cases} \tag{1.87}$$

then the BNT lensing kernel $\omega(z)_{\rm nulled}$ associated with our BNT convergence will be zero for $z < z_1$ confining the contributing lenses between z_1 and z_3.

The system given by equation (1.87) can now be easily solved which leads to

$$\begin{cases} p_1/p_3 = \dfrac{\chi(z_1)(\chi(z_2)-\chi(z_3))}{\chi(z_3)(\chi(z_1)-\chi(z_2))}, \\ p_2/p_3 = \dfrac{\chi(z_2)(\chi(z_3)-\chi(z_1))}{\chi(z_3)(\chi(z_1)-\chi(z_2))}, \end{cases} \tag{1.88}$$

p_3 thus being chosen as an arbitrary normalisation, usually 1.

1.9.3 Tomographic analysis

Assuming that enough galaxies can be observed in each bin, the usual strategy to maximise the information one can get out of a cosmological weak-lensing experiment is to allocate sources in redshift bins so as to gain information on the redshift evolution of clustering (Hu, 1999), and thus perform a full tomographic analysis. As far as the BNT transform is concerned, it allows one to go one step further following this idea by disentangling and sorting out (or decorrelating) physical scales probed by gravitational lensing. Indeed, noticing that the BNT transform applied to 3 successive source planes leaves the lensing kernel non zero only between the last 2, applying this transform iteratively to a set of n source planes will leave the BNT lensing kernels only overlapping two by two, which as a consequence, since lenses at each redshift are considered independent, decorrelates* for a large part the newly constructed nulled lensing fields. All that remains is to decide the fate of the first two source planes, which cannot be fully submitted to the nulling procedure. This seems rather logical in a sense, since the ultimate goal of the BNT transform is not so much to discard the information coming from the foreground lenses but to sort out the dynamical scales from which the signal is coming. The first two bins are thus (almost) left untouched. As a result and as will be explicit in the next paragraph, the BNT transform can be thought of as an invertible linear application applied to the set of observed maps and in that sense do not amount to any loss in the information content of the maps. Moreover, the fact that nulling is simultaneously ensured at all scales and in all regimes of gravitational instabilities potentially offers a way to use observations at small scales, even in regimes where predictions from theory or numerical experiments are at best approximate, either because of poor numerical resolution or unknown physical effects such as the impact of baryon physics. Indeed, since the weights p_i only depend on cosmological parameters, then an inexact estimation of those parameters would lead to an inexact nulling, demonstrated in data by a non-zero correlation between non-overlapping BNT bins.

Formally, I define the BNT transform as a linear transformation M applied to the set of lensing kernels $\omega_i \equiv \omega(z, z_i)$, giving rise to a new set of re-weighted kernels

$$\omega^i_{\rm null} = M^{ij}\omega_j. \tag{1.89}$$

All non-zero coefficients are given by

$$\begin{aligned} M^{i,i} &= 1, \quad M^{2,1} = -1 \\ M^{i,i-2} &= \frac{\chi_{i-2}(\chi_{i-1} - \chi_i)}{\chi_i(\chi_{i-2} - \chi_{i-1})}, \\ M^{i,i-1} &= \frac{\chi_{i-1}(\chi_i - \chi_{i-2})}{\chi_i(\chi_{i-2} - \chi_{i-1})}. \end{aligned} \tag{1.90}$$

Let me now give an example. I will consider the 8 first source planes of a Euclid-like experiment† and these are located at $z_s = \{0.209, 0.489, 0.619, 0.7335, 0.8445, 0.9595, 1.087, 1.2395, 1.45, 2.038\}$. The associated matrix M is then given by

*One way to say it is that it block-diagonalises their correlation matrix.

†These are actually the mean redshift of the first 8 equally-populated redshift bins of the Euclid experiment.

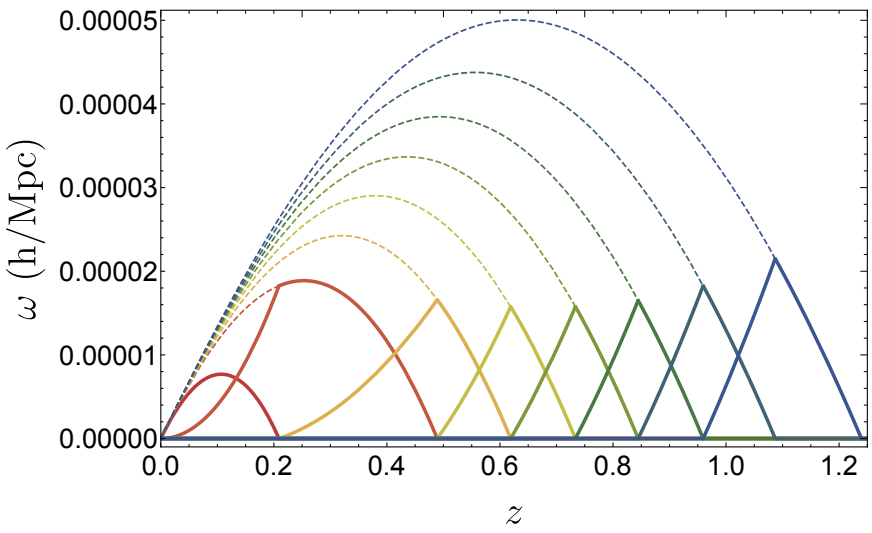


Figure 1.12: The radial selection functions, prior to nulling (dashed lines) and after nulling (thick lines). The source redshifts are taken as the mean redshift of the first 8 equally populated redshift bins of the Euclid experiment. This figure was published in **Barthelemy** et al. (2021a).

$$M = \begin{pmatrix} 1 & & & & & & \\ -1 & 1 & & & & & \\ 0.15 & -1.15 & 1 & & & 0 & \\ & 0.58 & -1.58 & 1 & & & \\ & & 0.71 & -1.71 & 1 & & \\ & & & 0.79 & -1.79 & 1 & \\ & 0 & & & 0.86 & -1.86 & 1 \\ & & & & & 0.92 & -1.92 \quad 1 \end{pmatrix}, \tag{1.91}$$

and I display in figure 1.12 the resulting BNT lensing kernels. The non-loss of information is here exemplified by the fact that no redshift ranges are left uncovered and the correlation structure of the constructed BNT lensing maps is even more apparent.

Finally, I would like to emphasise that though taking into account additional contribution to the signal such as shape noise does not in principle change anything to the full-joint analysis of all newly constructed nulled bins, the fact that each constructed BNT map results from the sum of 3 "normal" maps increases the noise terms that are added on top of the BNT map. This, plus the fact that the lensing signal is fainter within each BNT map, a consequence of narrowing the redshift range of the lensing kernel, renders it very hard to realistically exploit single BNT maps by themselves. However, since the transformation at the level of the full tomographic analysis does not present any loss in information, the way to go could be towards the joint study of BNT kernels to minimise this noise increase while still getting out the advantage of more localised lenses. This idea plus a more quantitative evaluation of noise, particularly shape noise, on single BNT bins will be studied in section 4.5 covering the modelling of the full aperture mass probability distribution functions. The formalism to then compute the full joint-PDF of nulled bins will be presented in section 4.6.

1.9.4 Source distributions

Let me now move on to the case of a distribution of sources $n(z)$ rather than discrete source planes. Still in a tomographic setting and for a distribution of sources throughout the survey of $n(z)$, the lensing kernel leading to the observed convergence in source bin i delimited by $z_{\mathrm{min},i}$ and $z_{\mathrm{max},i}$ is

$$\tilde{\omega}_i(z) = \int_{z \geq z_{\mathrm{min},i}}^{z_{\mathrm{max},i}} \frac{\mathrm{d}z_s}{H(z_s)} \frac{n(z_s)}{n_i^{(0)}} \omega(z, z_s), \tag{1.92}$$

where

$$n_i^{(0)} = \int_{z_{\mathrm{min},i}}^{z_{\mathrm{max},i}} \mathrm{d}z\, n(z). \tag{1.93}$$

As such and following the exact same derivation steps than for the discrete source planes case, a linear combination of the convergence field in 3 different redshift source bins would lead to a nulling of the lensing kernel between the observer and the lower boundary of the first source bin if the conditions

$$\begin{cases} \sum_{i=1}^{3} p_i = 0, \\ \sum_{i=1}^{3} p_i\, n_i^{(1)} = 0 \end{cases} \tag{1.94}$$

are met with

$$n_i^{(1)} = \int_{z_{\mathrm{min},i}}^{z_{\mathrm{max},i}} \mathrm{d}z\, \frac{n(z)}{n_i^{(0)} \chi(z)}. \tag{1.95}$$

As in the discrete source plane case, the system (1.94) is easily solved and the application to the first 8 source bins of the Euclid experiment this times leads to figure 1.13. Notice how the correlation structure evolves a bit switching from discrete source planes to extended source bins, the BNT kernels are now overlapping 3 by 3 instead of 2 by 2. Still, the property of sorting out physical scales is present, as is the property of decorrelating a large part of the signal coming from different source bins.

1.9.5 Limber approximation with the BNT transform

One of the main approximations that appears in every computation where relevant throughout this thesis is the Limber approximation. It consists of replacing spherical harmonics by Fourier transforms (the sky is considered flat), thus considering only small angle separations and this thus allows one to assume that only wave-modes perpendicular to the line of sight contribute to integrals along this line of sight, since $k_\perp \sim 1/\chi(z)\theta >> k_{||} \sim 1/\chi(z_s)$. This approximation is generally very precise and in any case well-tested, but one could wonder about its retained validity when narrowing quite heavily the lensing kernel as is the case with the BNT transform. To test this, I will rely on the formalism presented in LoVerde & Afshordi (2008), where building on a Taylor expansion of the spherical Bessel functions, they present all subsequent corrective terms to this Limber approximation on 2-point angular

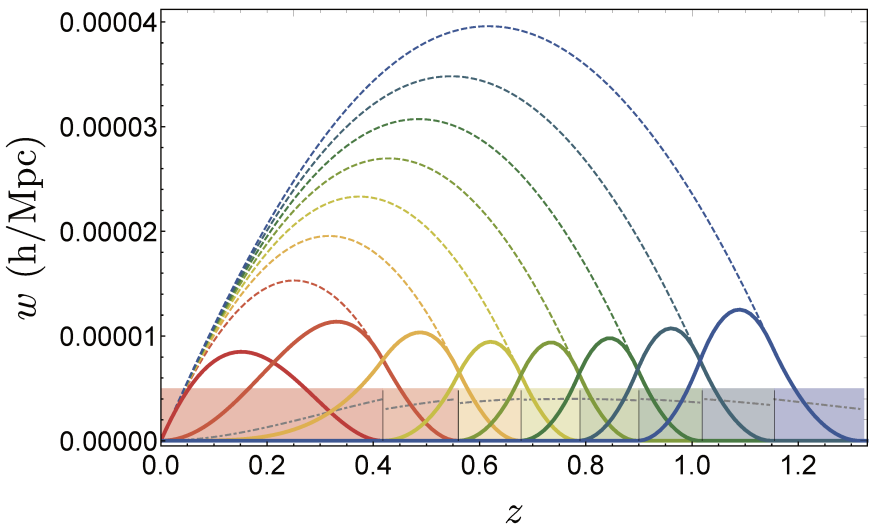


Figure 1.13: The realistic radial selection functions of the first 8 equally populated redshift bins (shown in coloured rectangles) of the Euclid experiment, prior to nulling (dashed lines) and after nulling (thick lines). The dot-dashed gray lines inside the coloured rectangles show the shape of the normalised distribution of sources inside each bin. This figure was published in **Barthelemy** et al. (2021a).

spectra. Their expression is

$$C_{AB}(\ell) = \int \frac{d\chi}{\chi^2} P_{AB}\left(\frac{\ell+1/2}{\chi}\right) f_A(\chi) f_B(\chi) \Bigg\{ 1 - \frac{1}{(\ell+1/2)^2}\left[\frac{r^2}{2}\left(\frac{f_A''(\chi)}{f_A(\chi)} + \frac{f_B''(\chi)}{f_B(\chi)}\right) + \frac{\chi^3}{6}\left(\frac{f_A'''(\chi)}{f_A(\chi)} + \frac{f_B'''(\chi)}{f_B(\chi)}\right)\right] + \mathcal{O}\left((\ell+1/2)^{-4}\right) \Bigg\} \quad (1.96)$$

where A and B are two 3-dimensional random fields, the matter density field or the gravitational potential for example, $f_{A,B}$ are the projection kernels and P_{AB} is the 3D cross power spectrum defined via $\langle A(\boldsymbol{k}_1)B(\boldsymbol{k}_2)\rangle = \delta^D(\boldsymbol{k}_1 - \boldsymbol{k}_2)P_{AB}(k_1)$. We first notice that indeed the zeroth order expression at large ℓ brings back the expression (1.38) we found for the convergence power spectrum. The first minor correction is to replace $\ell \to \ell + 1/2$, which only becomes relevant at low ℓ. The leading order correction then depends on the variation rate of the projection kernels, which indeed tends to be bigger for narrow lensing kernels. For sources located in places at $z_s = 1.2 - 1.4 - 1.6$ I then compute and display in figure 1.14 the convergence power spectrum given by its fully Limber approximated expression (1.38) and then the leading order correction that arises from equation (1.96). One can see that indeed the Limber approximation is valid for almost every angular scale except the very large ones, with the correction being several order of magnitudes below the zeroth order approximation. One could then apply this same technique to check the validity of the Limber approximation for every considered correlator. However, it is clear that the corrections will also depend directly on the variation rate of the lensing kernel, which is still going to be small enough and thus I will consider the Limber approximation to be valid throughout this thesis, be it for regular or nulled lensing quantities.

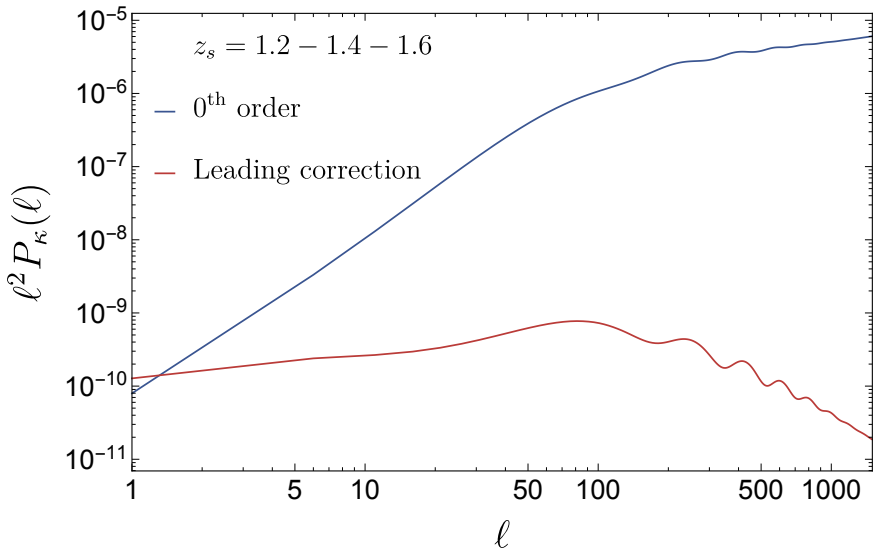


Figure 1.14: Illustration of the validity of the Limber approximation on the BNT convergence/shear power spectrum. The Limber approximation is still valid over a very large range of angular scales.

1.9.6 Impact of the additional lensing corrections on the BNT aperture mass skewness

To conclude this section on the BNT transform and given that one result of this thesis is its application to the convergence and aperture mass fields in the context of the computation of their one-point PDF, I now present analogous results to the post-Born corrections presented above but this time for the aperture mass (seen as a difference of the top-hat filtered convergence at two different scales (4.1)), and taking into account that I now also apply the BNT transform. These corrections, although small as I illustrate, unfortunately break the nulling property which now becomes inexact. The attentive reader may also realise that two additional corrections section 1.5.4 are missing from this one, the first correction coming from baryonic feedback that I do not know how to model and compute analytically but that we hope to avoid when performing theoretical computations in the context of the BNT transform, and the second one being intrinsic alignments whose impact on one-point statistics I have not – yet – tried to estimate. As for the remaining additional corrections that I mentioned, the impact of shape noise is straightforward to account for, as we treat it as an independent Gaussian noise on top of the convergence field, and the modelling of redshift space distortions is an ongoing work that already obtained some results but not advanced enough for them to be presented in this thesis. The impact of photometric redshifts on the exactness of the BNT transform is also part of an ongoing project.

Post-Born corrections: Extending the result we obtained on the convergence – negligible effect of the post-Born corrections on the convergence skewness – to the aperture mass boils down to computing the exact same post-Born corrections starting from equation (1.79), but this time taking into account that we implemented (summed maps at different redshift) the BNT transform and that the applied filter is now a difference of top hats rather than a

simple top-hat. We thus have

$$\left\langle M_{\rm ap}^3\right\rangle_{\rm corr} = \frac{-12}{(2\pi)^4}\int_0^z \frac{{\rm d}z'}{H(z')}\,\omega_{\rm (null)}(z',z)^2\int_0^{z'}\frac{{\rm d}z''}{H(z'')}\,\omega_{\rm (null)}(z'',z)\omega(z'',z')\int\frac{{\rm d}^2\boldsymbol{\ell}_1{\rm d}^2\boldsymbol{\ell}_2}{\left[\mathcal{D}(z')\mathcal{D}(z'')\right]^2}$$
$$P\left(\frac{\ell_1}{\mathcal{D}(z')},z'\right)P\left(\frac{\ell_2}{\mathcal{D}(z'')},z''\right)H(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2)W(\ell_1)W(\ell_2)W(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|), \quad (1.97)$$

where

$$H(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2) = \frac{\boldsymbol{\ell}_1\cdot\boldsymbol{\ell}_2}{\ell_2^2} + \frac{(\boldsymbol{\ell}_1\cdot\boldsymbol{\ell}_2)^2}{\ell_1^2\ell_2^2}, \quad (1.98)$$

and

$$W(l) = 2\frac{J_1(l\theta_2)}{l\theta_2} - 2\frac{J_1(l\theta_1)}{l\theta_1}, \quad (1.99)$$

is the $M_{\rm ap}$ window function in Fourier space and J_1 is the first order Bessel function of the first kind. The subscript (null) indicates where to input the nulled lensing kernel. Once again using the properties of the Bessel function (1.81) and (1.82) I find

$$\left\langle M_{\rm ap}^3\right\rangle_{\rm corr} = -\frac{12}{\pi^2\theta_1^3\theta_2^3}\int_0^z\frac{{\rm d}z'}{H(z')}\,\omega_{\rm (null)}(z',z)^2\int_0^{z'}\frac{{\rm d}z''}{H(z'')}\omega_{\rm (null)}(z'',z)\,\omega(z'',z')$$
$$\int\frac{{\rm d}\ell_1{\rm d}\ell_2}{\left[\mathcal{D}(z')\mathcal{D}(z'')\right]^2}P\left(\frac{\ell_1}{\mathcal{D}(z')},z'\right)P\left(\frac{\ell_2}{\mathcal{D}(z'')},z''\right)\frac{1}{\ell_2}\Bigg((\theta_1J_1(\ell_1\theta_2)-\theta_2J_1(\ell_1\theta_1))(\theta_1J_1(\ell_2\theta_2)$$
$$-\theta_2J_1(\ell_2\theta_1))(\theta_2(J_2(\ell_1\theta_1)-J_0(\ell_1\theta_1))J_1(\ell_2\theta_1)+\theta_1(J_0(\ell_1\theta_2)-J_2(\ell_1\theta_2))J_1(\ell_2\theta_2))\Bigg). \quad (1.100)$$

For a difference of top-hats with opening angles θ_1 and $\theta_2 = 2\theta_1$, I finally obtain the correction on the nulled $M_{\rm ap}$ skewness due to lens-lens coupling and geodesic deviation shown in figure 1.15 for source redshifts located at $z_s = 1.2, 1.4$ & 1.6 as a function of θ_1. As expected, the correction is shown to i) Gaussianise the field and ii) be very small – sub-percent – which is not surprising since reducing the lensing kernel diminishes the importance of couplings between lenses.

Reduced shear correction: I have already mentioned that assuming that the intrinsic ellipticity of galaxies has no preferred orientation, the observed ellipticity is an unbiased estimator of the reduced shear $g = \gamma/(1-\kappa)$ rather than the shear itself. This implies that rather than equation (1.42), the observed physical aperture mass is given by

$$\begin{aligned} M_{\rm ap}^g(\boldsymbol{\vartheta}) &= \int {\rm d}^2\boldsymbol{\vartheta}' Q_\theta\left(\vartheta'\right) g_t\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'\right) \\ &\approx M_{\rm ap}(\boldsymbol{\vartheta}) + \int {\rm d}^2\boldsymbol{\vartheta}' Q_\theta\left(\vartheta'\right)\gamma_t\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'\right)\kappa\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'\right) \\ &\approx M_{\rm ap}(\boldsymbol{\vartheta}) + \delta M_{\rm ap}(\boldsymbol{\vartheta}). \end{aligned} \quad (1.101)$$

I thus derive in this subsection the leading correction to the aperture mass skewness that accounts for the reduced shear. This is merely a perturbation theory calculation, which

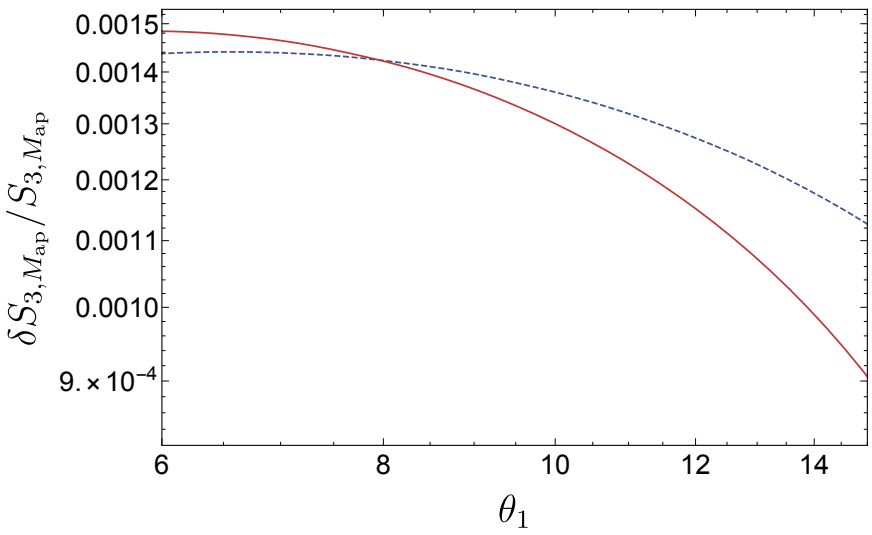


Figure 1.15: Impact of the leading order corrections to the BNT $M_{\rm ap}$ skewness induced by post-Born (blue) and reduced shear (red) terms at source redshifts $z_s = 1.2, 1.4$ & 1.6. The field is filtered by a difference of top-hats with opening angles θ_1 and $\theta_2 = 2\theta_1$ and we plot the evolution of the correction with respect to θ_1 in arcmin. Dashed lines indicate negative values. This figure was published in **Barthelemy** et al. (2021b).

still somewhat relies on the fact that the convergence κ is small and described by linear perturbation theory. Though this is rather straightforward and very useful for an estimation of the amplitude of the effect, this is not at the same level than what can be performed using large deviation theory to account for the effect as is shown without any projection effects for example in Reimberg & Bernardeau (2018).

First denoting the wave vector $\boldsymbol{\ell} = (\ell\cos\varphi_s, \ell\sin\varphi_s)$ conjugate of $\boldsymbol{\vartheta} = (\vartheta\cos\varphi, \vartheta\sin\varphi)$ and combining equations (1.36) and (1.41) I get the tangential shear γ_t as a function of the convergence κ

$$\gamma_t(\boldsymbol{\vartheta}) = -\int \frac{\mathrm{d}^2\boldsymbol{l}}{2\pi}\cos\left(2(\varphi-\varphi_s)\right)e^{i\boldsymbol{\ell}\cdot\boldsymbol{\vartheta}}\tilde{\kappa}(\boldsymbol{\ell}). \tag{1.102}$$

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

$$\left\langle (M_{\rm ap}^g)^3\right\rangle \approx \left\langle M_{\rm ap}^3\right\rangle + 3\left\langle M_{\rm ap}^2\,\delta M_{\rm ap}\right\rangle \tag{1.103}$$

where the reduced shear correction is thus written as

$$\begin{aligned}\left\langle M_{\rm ap}^2\,\delta M_{\rm ap}\right\rangle = -\int \mathrm{d}^2\boldsymbol{\vartheta}_1 U_\theta(\vartheta_1)\int\frac{\mathrm{d}^2\boldsymbol{\ell}_1}{2\pi}e^{i\boldsymbol{\ell}_1\cdot\boldsymbol{\vartheta}_1}\int\mathrm{d}^2\boldsymbol{\vartheta}_2 U_\theta(\vartheta_2)\int\frac{\mathrm{d}^2\boldsymbol{\ell}_2}{2\pi}e^{i\boldsymbol{\ell}_2\cdot\boldsymbol{\vartheta}_2}\int\mathrm{d}^2\boldsymbol{\vartheta}Q_\theta(\vartheta)\\ \int\frac{\mathrm{d}^2\boldsymbol{\ell}'}{2\pi}e^{i\boldsymbol{\ell}'\cdot\boldsymbol{\vartheta}}\int\frac{\mathrm{d}^2\boldsymbol{\ell}}{2\pi}e^{i\boldsymbol{\ell}\cdot\boldsymbol{\vartheta}}\cos\left(2(\varphi-\varphi_s)\right)\left\langle\tilde{\kappa}(\boldsymbol{\ell}_1)\tilde{\kappa}(\boldsymbol{\ell}_2)\tilde{\kappa}(\boldsymbol{\ell})\tilde{\kappa}(\boldsymbol{\ell}')\right\rangle.\end{aligned} \tag{1.104}$$

Again stopping at leading order, I only consider the linear evolution of the matter density fluctuation, which means that the κ field is Gaussian. This allows me to use Wick's theorem to compute the correlator in equation (1.104) which becomes

$$\begin{aligned}\left\langle\tilde{\kappa}(\boldsymbol{\ell}_1)\tilde{\kappa}(\boldsymbol{\ell}_2)\tilde{\kappa}(\boldsymbol{\ell})\tilde{\kappa}(\boldsymbol{\ell}')\right\rangle = P_\kappa(\ell_1)P_\kappa(\ell')\delta^D(\ell_1+\ell_2)\delta^D(\ell'+\ell)+P_\kappa(\ell_1)P_\kappa(\ell_2)\delta^D(\ell_1+\ell)\delta^D(\ell'+\ell_2)\\ + P_\kappa(\ell_1)P_\kappa(\ell_2)\delta^D(\ell_1+\ell')\delta^D(\ell+\ell_2).\end{aligned} \tag{1.105}$$

The first term will yield zero because of the integration of the cosine, and the two other terms yield the same contribution. Carrying out the integration other $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2$ is straightforward thanks to the presence of the Dirac delta functions and I arrive at

$$\left\langle M_{\rm ap}^2\,\delta M_{\rm ap}\right\rangle = -2\int \mathrm{d}^2\boldsymbol{\vartheta}_1 U_\theta(\vartheta_1)\int \mathrm{d}^2\boldsymbol{\vartheta}_2 U_\theta(\vartheta_2)\int \mathrm{d}^2\boldsymbol{\vartheta} Q_\theta(\vartheta)\int \frac{\mathrm{d}^2\boldsymbol{\ell}}{(2\pi)^2}e^{i\boldsymbol{\ell}\cdot(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_1)}P_\kappa(\ell)$$
$$\cos\left(2(\varphi-\varphi_s)\right)\int \frac{\mathrm{d}^2\boldsymbol{\ell}'}{(2\pi)^2}e^{i\boldsymbol{\ell}'\cdot(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_2)}P_\kappa(\ell'). \quad (1.106)$$

The integration over $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$ leads to the Fourier expression of the aperture mass filter and all that remains is to perform the angular integrations over $\boldsymbol{\vartheta}$, $\boldsymbol{\ell}$ and $\boldsymbol{\ell}'$. These yield expressions that correspond to the integral definition of the Bessel functions of the first kind

$$J_n(|x|) = \frac{1}{2\pi(i)^n}\int_{-\pi}^{\pi}\mathrm{d}\varphi\, e^{i(x\cos(\varphi)-n\varphi)},\ n\in\mathbb{N},\ x\in\mathbb{R}. \quad (1.107)$$

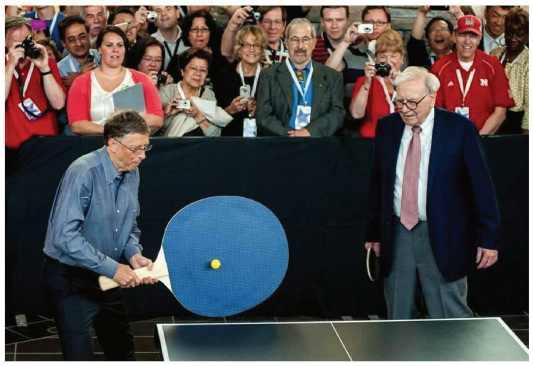
The expression for the reduced shear correction to the skewness of the aperture mass is thus

$$\left\langle (M_{\rm ap}^g)^3\right\rangle_{\rm corr} = \frac{3}{\pi}\int \mathrm{d}\ell\ell C_\ell^\kappa(\ell)W(\ell)\int \mathrm{d}\ell'\ell' C_\ell^\kappa(\ell')W(\ell')\int \mathrm{d}\vartheta\vartheta Q_\theta(\vartheta)J_0(\vartheta\ell')J_2(\vartheta\ell). \quad (1.108)$$

I now evaluate this correction in the same configuration as the post-Born corrections, that is nulling with evolving opening angles, and plot the resulting correction in figure 1.15. Finally, note that in the case of the BNT transform, the procedure is only exact for the convergence and shear fields and thus another term accounting for this issue should also be present when introducing the reduced shear correction. Taking into account all corrections that arise because of inaccuracies of the procedure in realistic settings is part of an ongoing work.

Magnification bias correction: As shown in equation (1.48) and the corresponding section 1.5.4, the leading order correction to the skewness coming from the magnification bias correction is exactly the same as in the case of the reduced shear in equation (1.108) up to a factor $(5s-2)$. Indeed, remember that the source-lens clustering term will yield a zero contribution in any correlator if the lenses and sources do not overlap which is the case when considering source planes or very narrow redshift bins and more importantly when applying the BNT transform to any tomographic bins.

Using the words
«Large deviation theory»
in cosmology papers



2

Large deviation theory for the cosmologist

Contents

2.1. Introduction and brief history

The mathematical theory of large deviations is a rich area of probability theory that concerns itself with the asymptotic behaviour of the far tails of probability distributions. Roughly speaking, large deviation theory describes the exponential decline of the probability measures of certain kinds of rare events. For our purpose, it is best to see it as an extension of the central limit theorem, itself seen as an extension of the law of large numbers. It is also a rigorous formalism for the application of the saddle point approximation and indeed the first basic ideas of the theory can be traced back all the way to Pierre-Simon de Laplace at the

beginning of the 19$^{\text{th}}$ century, with his derivation of a more general central limit theorem made by the introduction of characteristic functions. A first attempt at a formal formulation of the theory was however performed much later, in the 1930s, by Harald Cramér, a Swedish mathematician who was concerned with the sum of independent and identically distributed (IID) random variables and first expressed rate functions (which I define below) as power series. Starting from there, many mathematicians made several important contributions to the theory, the most well-known being Sanov's theorem in 1957, Varadhan in the 1970s who first wrote a unified formalisation of large deviation theory and won the 2007 Abel prize for it, and finally – for what is of interest to us – the theorem by Gärtner and Ellis at the beginning of the 1980s. The field of large deviations is still active nowadays, with many efforts from applied mathematicians to describe physical systems out of equilibrium, disordered systems, chaotic and even quantum systems, and more generally subjects within the field of statistical physics where the number of interacting bodies can be modelled to go to infinity. Indeed, from the physicist's point of view, it would seem that though large deviations are not commonly taught in physics courses, statistical physicists have been using it (unknowingly for the most part before it even became a mathematical theory) for a hundred years and key concepts such as entropy or free energy are cast in a new light when respectively seen as a rate function and the associated scaled cumulant generating function. A very complete book that provides a rigorous approach to both large deviations and applications to statistical mechanics, though somewhat difficult to read, was written by Richard Ellis (1985). A simpler but very useful review, this time directly aimed towards physicists discovering the subject, can be found in Touchette (2009).

This chapter aims at giving a theoretical astrophysicist's point of view on the theory of large deviations in the cosmological context. Thus both the formalism and the existing theorems will be given in a somewhat less rigorous manner that what could be found in a large deviations textbook but with the advantage of being much more accessible to astrophysicists. The aim of this chapter is to provide intuition on how those concepts can give us profound insights on the dynamics of large-scale structure and how they fit very well within the more commonly used perturbative frameworks. I will start with a very simple but hopefully helpful example of 'large deviations behaviour' with the sample mean of independent and identically distributed random variables before moving to a more formal definition of the objects and useful theorems. Though I do not aim at reproducing rigorous proofs, I will also try to provide insight on how they work when possible. This part is inspired from the presentation found in Touchette (2009). I will make a small detour towards statistical physics and finally come back to cosmology and present how this framework is used to compute various statistics of the matter density field.

2.2. An introductory example: Sample mean of IID random variables

The well-known (in astrophysics) Poisson distribution deals with the number of random independent occurrences in a fixed period of time, for example if a photon from a source has a chance to be emitted with the same probability at each instant regardless of previous emissions, then the number of photons hitting a detector per fixed exposition time will follow

the Poisson distribution. This needs to be taken into account for example to measure the source luminous flux and is commonly known as shot noise. In a case more related to the works in this thesis, the Poisson distribution can also model the scatter of the luminous tracer density around its mean value $\langle\delta_{\rm tracer}|\delta_{\rm matter}\rangle$ although more elaborate models seem to be much more reliable (Friedrich et al., 2021).

A distribution closely related to the Poisson distribution is the exponential distribution, which deals with the time between occurrences of successive Poisson events as time flows by, for example the time between phone calls one may receive during a work day not counting breaks, or the time between looks at my watch to see if it is time to leave the office on a sunny Friday afternoon. A (strictly positive) continuous random variable X of mean μ following an exponential distribution admits a PDF of the form

$$\mathcal{P}_X(x) = \frac{1}{\mu}\exp\left[-\frac{x}{\mu}\right]. \tag{2.1}$$

Now let me consider n independent random variables X_i with $i \in [\![1, n]\!]$ following (2.1) and let me define their sample mean S_n as

$$S_n = \frac{1}{n}\sum_i^n X_i. \tag{2.2}$$

There are several methods to then compute the induced PDF of S_n. We could for example rely on the existence and knowledge of the cumulant generating function (see appendix A.1.2) of exponential distributions and sum them for each X_i since they are independent, or we could also straightforwardly directly compute their successive convolution and find the appropriate recursion rule to obtain the resulting PDF. The first method is the one I will use later in the projection formula (3.4) so here I use the second one and obtain the Erlong distribution of parameter n/μ.

$$\mathcal{P}_{S_n}(s) = \left(\frac{n}{\mu}\right)^n x^{n-1}\frac{\exp\left[-nx/\mu\right]}{(n-1)!}. \tag{2.3}$$

I plot the resulting distribution in figure 2.1 as a function of n for the case of $\mu = 1$. As expected from the law of large numbers, as n grows the distribution peaks more and more around the mean value $\mu = 1$ to the point of approaching a Dirac delta function as n goes to infinity. An extension of this is found in the central limit theorem, which states that the fluctuations around the mean as n tends to infinity are Gaussian distributed, which can be seen in the figure by looking at how the distribution is more and more symmetric around $s = \mu = 1$. However, further away in the tails of the distribution, it turns out that it is not Gaussian.

In order to see this, let me note that the Erlong distribution of parameter n/μ has a particular asymptotic behaviour as n grows so that we can re-write it as

$$\mathcal{P}_{S_n}(s) \asymp \exp[-n\psi(s)] \tag{2.4}$$

with

$$\psi(s) \equiv \lim_{n\to+\infty} -\frac{1}{n}\log(\mathcal{P}_{S_n}(s)) = \frac{s}{\mu} - 1 - \log\left(\frac{s}{\mu}\right), \tag{2.5}$$

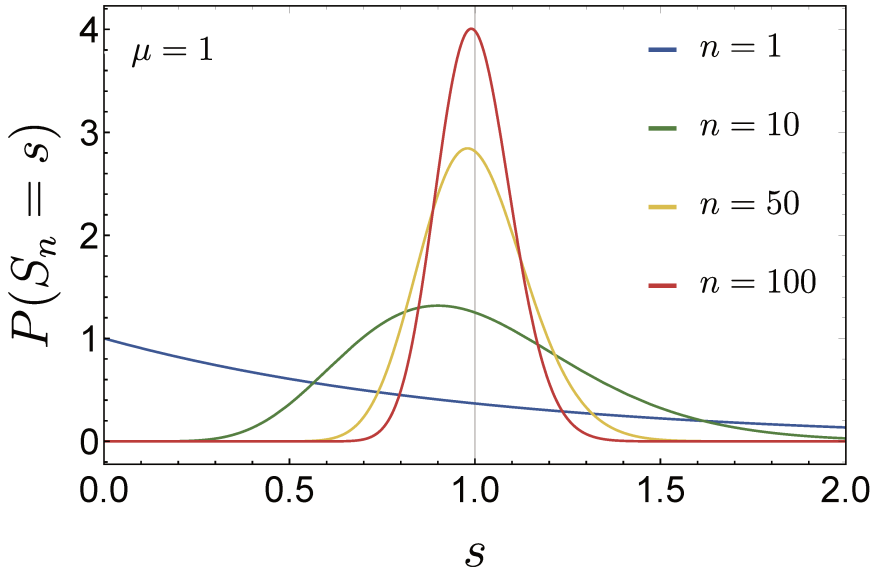


Figure 2.1: Distribution of the sample mean of n IID random variables following an exponential distribution of parameter $\mu = 1$. As n grows, the distribution becomes more and more peaked around the mean value $\mu = 1$ approaching a dirac $\delta_D(s - \mu)$ thus illustrating the law of large numbers. As n continues to grow, the distribution increasingly resembles a Gaussian distribution, thus illustrating the central limit theorem. However, looking at the asymptotic expression (2.5), one finds that though a Taylor expansion of ψ around μ would lead to a Gaussian, further away from the fluctuations around the mean, in the rare events tails, the distribution is not exactly Gaussian. These are the *large deviations* that the theory aims at studying.

and where the precise definition of the $\asymp$ symbol will be given by the large deviation principle in the next section 2.3 but illustrates the asymptotic behaviour of $\mathcal{P}_{S_n}(s)$, its dominant behaviour since other terms are sub-exponential in n. We thus observe that as n goes to infinity, the tails of the PDF decay exponentially at a rate given by $\psi(s)$ that is thus called the rate function. Additionally, and to further illustrate this, I plot in figure 2.2 the convergence of $-\frac{1}{n}\log(\mathcal{P}_{S_n}(s))$ to $\psi(s)$ as n grows and one can notice that, interestingly, the convergence to $\psi(s)$ can be quite fast.

That is thus basically what large deviation theory deals with: the exponential rate at which probabilities of rare events – rare because away from the most probable value – decay as a natural parameter of the problem varies, here the number of IID random variables in the sum. Moreover, the general exponential form $\exp(n\psi(s))$ that we found for our previous sample mean, and which is the founding result of large deviation theory, arises in the context of many stochastic processes, not just IID sample means, which is the main reason why a whole theory can be built around this idea.

2.3. Large deviation theory: Mathematical foundations

The large deviation principle: The rigorous, mathematical definition of the large deviation principle involves concepts of topology and measure theory that are too technical to be presented here and also not exactly useful for the cosmological context. They can however be found in the aforementioned book by Richard Ellis (1985) and the appendices of

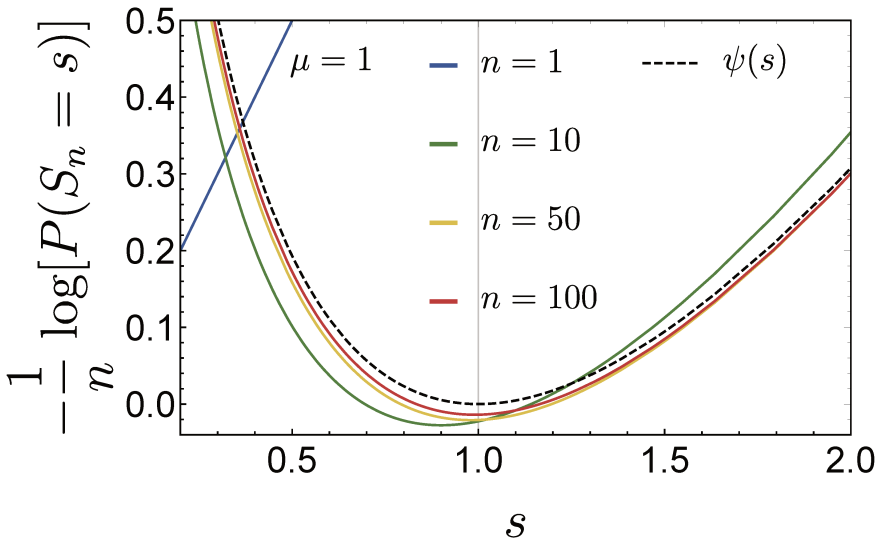


Figure 2.2: Convergence of $-\frac{1}{n}\log(\mathcal{P}_{S_n}(s))$ to the rate function $\psi(s)$ as n grows. The convergence is found to be quite fast which can make the approximation of the sample mean distribution by the exponential of its associated rate function $\psi(s)$ relatively precise.

the review by Touchette (2009). For simplicity, I will restrict myself to the case of continuous random variables X that admit a PDF $\mathcal{P}_X(x)$ and say that X satisfies a large deviation principle (LDP) if the following limit exists:

$$\psi(x) \equiv \lim_{n\to+\infty} -\frac{1}{n}\log(\mathcal{P}_X(x)) \tag{2.6}$$

and gives rise to the rate function $\psi(x)$ which is not everywhere zero*.

THE GÄRTNER-ELLIS THEOREM: One of the main results of large deviation theory is based on the computation of the following function, the scaled cumulant generating function (SCGF) defined as

$$\varphi(\lambda) = \lim_{n\to\infty} \frac{1}{n}\log\left[E(e^{n\lambda X})\right] \tag{2.7}$$

and named like this since it is indeed a re-scaling of the usual cumulant generating function (CGF) $\phi(\lambda)$ defined in appendix A.1.2. Here λ is a real parameter, $E(.)$ denotes the expectation value, and X is a collection of arbitrary random variable indexed by n. The theorem then states that if $\varphi(\lambda)$ exists and is differentiable, then X satisfies a large deviation

*Note again that this definition of the large deviation principle is a simplification of the rigorous definition used in mathematics, due to the mathematician Srinivas Varadhan, is expressed in terms of probability measures of certain sets rather than in terms of probability densities, and involves upper and lower bounds (and thus upper and lower rate functions) on these probabilities rather than a simple limit (the upper and lower rate functions are then considered equal). The mathematical definition also applies a priori to any random variables, not just continuous ones who admit a PDF. The definition I use is much closer to the form present in Cramér's theorem in the 1930's exclusively for sums of IID random variables as in the previous example I gave.

principle with rate function ψ given by the *Legendre-Fenchel* transform of the SCGF

$$\psi(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \varphi(\lambda)\}. \tag{2.8}$$

Note that since $\varphi(\lambda)$ is always convex, the Legendre-Fenchel transform is reduced to a more simple Legendre transform expressed by removing the supremum and having λ as a function of x by solving the condition $x = \mathrm{d}\varphi/\mathrm{d}\lambda$ (given by definition of the Legendre transform).

A heuristic proof of the second part of the Gärtner-Ellis theorem (the expression of the rate function) can be given by making use of the saddle point approximation, which is justified in the context of large deviation theory. If I already assume that X satisfies a large deviation principle with rate function ψ then we have

$$\left\langle e^{n\lambda X}\right\rangle \asymp \int \mathrm{d}x\, e^{n(\lambda x-\psi(x))}, \tag{2.9}$$

which can be computed quite naturally using a saddle-point approximation because the error associated with it is of the same order as the error associated with the large deviation approximation. As a consequence assuming that the maximum of $\lambda x - \psi(x)$ exists and is unique, we find

$$\left\langle e^{n\lambda X}\right\rangle \asymp \exp\left(n\sup_x\{\lambda x - \varphi(\lambda)\}\right). \tag{2.10}$$

From this we have that the SCGF is the Legendre-Fenchel transform of the rate function and is thus convex ($\varphi''(\lambda) \geq 0$). Then, provided that φ is differentiable, the Legendre-Fenchel transform is self-dual and we arrive at (2.8). Note that though the SCGF can always be written as the Legendre-Fenchel transform of the rate function, thus always being convex, the differentiability condition in the Gärtner-Ellis theorem is very important since the rate function itself might not be convex, in which case the Legendre-Fenchel transform of the SCGF only gives the convex envelope of the rate function. Non-convex rate-functions will not be encountered in our case but usually correspond to multi-modal PDFs (multi-minima of the rate functions give rise to different peaks in the PDF) and as such we will always assume that the convex envelopes (the linear parts connecting the convex bits) of rate functions are the rate functions. Indeed, in the cosmological context we will assume that generating functions can be expanded in power-series whose coefficients will be linked to results from perturbation theory, which thus renders our SCGFs differentiable.

Varadhan's theorem: We just showed that thanks to the saddle point approximation we have

$$\varphi(\lambda) = \sup_x\{\lambda x - \psi(x)\}. \tag{2.11}$$

Then replacing λx by an arbitrary continuous function f of X gives Varadhan's theorem, stated as

$$\varphi[f] = \lim_{n\to\infty}\frac{1}{n}\log\left[E(e^{nf(X)})\right] = \sup_x\{f(x) - \psi(x)\}. \tag{2.12}$$

The function $\varphi[f]$ is thus defined as a functional of f, and though Varadhan's theorem is a consequence of the saddle point approximation when X is a real random variable, for

other types of random variables, such as random functions, Varadhan's theorem still applies, and thus extends the saddle point approximation to these random variables. The theorem also holds when $f(x)\psi(x)$ has more than one maximum, that is when the integral defining $E(e^{nf(X)})$ has more than one saddle point. This theorem is perhaps the most important one for the application of large deviation theory in the cosmological context, as getting the rate function for the matter density field is indeed relatively easy but what we are looking for are the cumulants of the field to then re-construct the PDF.

PROPERTIES OF φ AND ψ: The first property that is worth mentioning is that both the rate function and the SCGF are unique. Note however that different rate functions can lead to the same convex envelope. From the definition of the SCGF, we can extract several properties regarding its link to the cumulants of the random variable X. First we get that since the PDF is normalised to one, $\varphi(0) = 0$. Then we have

$$\varphi'(0) = \lim_{n\to\infty} \left.\frac{\left\langle X \mathrm{e}^{n\lambda X}\right\rangle}{\left\langle \mathrm{e}^{n\lambda X}\right\rangle}\right|_{k=0} = \lim_{n\to\infty} \langle X\rangle \,, \tag{2.13}$$

provided that the SCGF is indeed differentiable at zero. We similarly have

$$\varphi''(0) = \lim_{n\to\infty} n\left(\left\langle X^2\right\rangle - \langle X\rangle^2\right) = \lim_{n\to\infty} n \operatorname{var}(X) \tag{2.14}$$

and more generally

$$\varphi^{(p)}(0) = \lim_{n\to\infty} n^{p-1} \left\langle X^p\right\rangle_c \tag{2.15}$$

where the subscript c indicates the connected part of the moments (the cumulants). This property gives further credit to the "scaled" in SCGF. I have already mentioned that the SCGF is also always convex.

Rate functions are always positive, as can be seen from Varadhan's theorem and the fact that $\varphi(0) = 0$

$$\varphi(0) = \sup_x\{-\psi(x)\} = -\inf_x \psi(x) = 0. \tag{2.16}$$

LAW OF LARGE NUMBERS AND CENTRAL LIMIT THEOREM: If ψ admits a unique and global minimum zero x^* then it corresponds to the only value at which the PDF does not decay exponentially, and so around which the PDF gets more and more concentrated as $n \to \infty$ and we have

$$x^* = \varphi'(0) = \lim_{n\to\infty} \langle X\rangle \,. \tag{2.17}$$

This concentration notably means that

$$\lim_{n\to\infty} P\left(X \in [x^*, x^* + \mathrm{d}x]\right) = 1 \tag{2.18}$$

which is nothing but a formulation of the law of large numbers, where we also have the notable advantage of knowing the rate at which X converges in probability to its mean. In general, the existence of a law of large numbers for a random variable X is a good sign that a large deviation principle holds for X and this can be used as a point of departure for

deriving large deviation principles, such as how the matter density field converges to its mean value at very high redshifts and/or very large scales. However, ψ might admit more than one global minimum in which case the law of large numbers does not hold. Several global minima would then yield typical values of X just as in the case of one minimum while local minima would yield values of X for which the PDF is locally but not globally maximum. In statistical physics, the global minimum might be referred to as an "equilibrium state" while local minima give "metastable" values of the studied field.

Again admitting that ψ has a unique and global minimum zero x^* and is twice differentiable at x^* we can Taylor expand it as

$$\psi(x) \approx \frac{1}{2}\psi''(x^*)(x - x^*)^2, \tag{2.19}$$

which leads to a Gaussian approximation of the PDF. From the theorem of Berry-Esseen, the Gaussian approximation displayed above can be shown to be accurate for values of X around x^* of order $n^{-1/2}$. This means that for small fluctuations around the mean the central limit theorem yields essentially the same information as large deviation theory, while a large deviation away from the mean is a value of X for which the rate function departs significantly from its quadratic Taylor expansion. We can thus see large deviation theory as a generalisation of the central limit theorem*.

Contraction principle: Up to now I have summarised two main results of large deviation theory, namely the Gärtner-Ellis and the Varadhan's theorems, and there is only one remaining key result that I would like to introduce, the contraction principle. Let us again consider a random variable X that admits a large deviation principle with rate function $\psi_X(x)$ and another random variable Y linked to X via the continuous mapping $Y = h(X)$. We then might ask whether Y satisfies a large deviation principle itself, and if so, with which rate function. The answer to the first question is yes, Y admits a large deviation principle, and we can easily build intuition as to why and how to compute its rate function, again making use of a saddle point approximation in the large deviation limit. Indeed we have

$$P(Y \in \mathrm{d}y) = \int_{x:h(x)=y} P(X \in \mathrm{d}x), \tag{2.20}$$

and expressing the PDF of X in terms of its rate function in the large deviation limit and taking the saddle point approximation leads to

$$P(Y \in \mathrm{d}y) \asymp \mathrm{d}x \exp\left(-n \inf_{x:h(x)=y} \psi_X(x)\right). \tag{2.21}$$

This shows that indeed Y satisfies a large deviation principle with rate function

$$\psi_Y(y) = \inf_{x:h(x)=y} \psi_X(x). \tag{2.22}$$

*Note that there are also random variables which admit a central limit theorem without their rate function being expandable quadratically. See for example the sample mean of double-sided Pareto random variables presented as example III.4 in Touchette (2009).

This formula is thus called the contraction principle because h can be many-to-one in which case we are *contracting* information about the rate function of one random variable down to the other. In physical terms, this states that an improbable fluctuation of Y is brought about by the most probable of all improbable fluctuations of X. If h were a bijective function we would just have $\psi_Y(y) = \psi_X(h^{-1}(x))$. In the language of path integrals, the contraction principle just worries about the most probable, the classical, path/mapping between the two random variables.

2.4. Entropy, free energy and Legendre transforms in statistical mechanics

Before moving specifically to the case of the matter density field and how it can be modelled with the help of large deviation theory, since it mostly finds its application in the realm of statistical physics, let me very briefly make the link between the different objects I introduced in the previous section and the concepts of entropy, free energy and why a Legendre transform links the two. This section is however very basic and intended only to give a taste of what things might look like for more complex examples. It is strongly inspired by the more involved examples given in Touchette (2009) whose purpose is much more oriented towards the link between large deviation theory and statistical physics.

Let me consider a set of n particles interacting through some sort of potential. These physical interactions are determined by a Hamiltonian (an energy function, see any analytical mechanics, statistical physics or quantum theory textbook for more details) H_n which then allows one to define the mean energy per particle $h_n = H_n/n$. The joint state of the particles is denoted by a micro-state $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ of n variables with ω_i denoting the state of the i^{th} particle and which gives the complete description of the state of each particle at the microscopic level. The space of all microstates Λ_n is then the n-fold product Λ^n of the one-particle state space Λ. Finally, we model the microstate ω as a random variable distributed according to a prior probability measure $P(\mathrm{d}\omega)$ on Λ_n, which is determined by physical considerations. In most cases, Liouville's theorem dictates that it is the uniform measure $P(\mathrm{d}\omega) = \mathrm{d}\omega/|\Lambda|^n$, where $|\Lambda|^n$, a constant, is the volume of Λ_n. Let us then study the large deviations of the mean energy $h_n(\omega)$ with respect to the prior $P(\mathrm{d}\omega)$. We have

$$P(h_n \in \mathrm{d}u) = \int_{\{\omega \in \Lambda_n : h_n(\omega) \in \mathrm{d}u\}} P(\mathrm{d}\omega), \tag{2.23}$$

which, since the prior measure $P(h_n \in \mathrm{d}u)$ is nothing but the Lebesgue measure $\mathrm{d}\omega$ up to a constant, is proportional to the volume

$$\Omega(h_n \in \mathrm{d}u) = \int_{\{\omega \in \Lambda_n : h_n(\omega) \in \mathrm{d}u\}} \mathrm{d}\omega. \tag{2.24}$$

If we thus now write the rate function $\psi(u)$ of $P(h_n \in \mathrm{d}u)$ we find that

$$\psi(u) = \ln \Omega - s(u) \tag{2.25}$$

with

$$s(u) = \lim_{n \to \infty} \frac{1}{n} \ln \Omega \left(h_n \in \mathrm{d}u \right), \tag{2.26}$$

the usual *microcanonical entropy* with the Boltzmann constant k_B set to 1. Thus if the rate function $\psi(u)$ exists, it is the entropy up to an additive constant and a minus sign.

Moreover, since $P(h_n \in \mathrm{d}u) \propto \Omega\left(h_n \in \mathrm{d}u\right)$, we have the scaled cumulant generating function $\varphi(\lambda)$ satisfying

$$\varphi(\lambda) = -\left.F(\beta)\right|_{\beta=-\lambda} - \ln|\Lambda|, \tag{2.27}$$

with

$$F(\beta) = \lim_{n\to\infty} -\frac{1}{n}\ln Z_n(\beta), \tag{2.28}$$

and

$$Z_n(\beta) = \int_{\Lambda_n} \mathrm{e}^{-n\beta h_n(\omega)}\mathrm{d}\omega = \int_{\Lambda_n} \mathrm{e}^{-\beta H_n(\omega)}\mathrm{d}\omega. \tag{2.29}$$

We thus recognise Z_n as the *partition function* associated with H_n, which makes $F(\beta)$ the *free energy function* as well as being $\varphi(\lambda)$ up to constant and a change of variable.

Having established the link between the entropy and the rate function, and the free energy and the SCGF, it is now much easier to explain why the two are linked by a Legendre transform in statistical mechanics textbooks. Indeed, the Varadhan and Gärtner-Ellis theorems allow us to reach the same result through pure mathematical arguments and even extending what is traditionally stated in statistical mechanics by giving precise conditions on F and s for them to be Legendre conjugates, or related by a Legendre-Fenchel transform, or when this is not the case for example when the conditions of the Gärtner-Ellis theorem are not satisfied. Indeed, we for example never recover the full non-concave entropy function from the free energy if it is not differentiable.

2.5. Large deviations of the cosmic matter density field

As is the case in the field of statistical physics, many equations that one arrives at when identifying a large deviation principle at play for the matter density field in cosmology were known long before their re-phrasing in the language of large deviation theory. The review by Bernardeau et al. (2002) for example treats some of those that I will write down in this section and some behaviour (for example the exponential decay of the PDF at the critical density, which I will define and explain in the following) were known as early as in Balian & Schaeffer (1989). However, the full understanding of these equations in the context of large deviation theory is fairly recent and is done for the first time in Bernardeau & Reimberg (2016). In this section and for reasons that will very soon become apparent I will restrict myself to the case of the matter density field filtered within top-hat (spherical) windows, but I will also present later how these results can then be extended to more general filters in section 4.4.

2.5.1 Establishing a large deviation principle and computing the rate function

Historically, this field started in 1992 in Bernardeau (1992) where the vertex generating function (see appendix A.3.2 for a definition) was shown to be given by spherical collapse dynamics. These vertices give rise not to the cumulants $\langle\delta^p\rangle_c$ of the matter density field but the scaled cumulants given by $S_p = \langle\delta^p\rangle_c/\sigma_\delta^{2(p-1)}$ with $p \geq 2$ at tree-order (leading order)

in standard Eulerian perturbation theory, which are then simple numbers* with no time or scale dependence (only with no smoothing applied to the density field). In this perturbative approach, the small parameter is the matter density contrast itself, which makes its variance small and thus the S_p parameters obtained through the spherical collapse dynamics are accurate for describing the non-linear evolution of the matter density field only when the variance is taken in its limit to zero. In the zero variance limit, that corresponds to either or both very high redshift and very large scales, and the density field peaks towards its most probable value, which is thus the mean density of the universe (homogeneous background with no fluctuations) with a behaviour close to the law of large numbers. Moreover, the small fluctuations around this mean value are Gaussian† as a result of linear perturbation theory, and as can be seen from the expression of the PDF in terms of its cumulants of high-order p that go to zero as fast as a power $p-1$ of the speed at which the variance itself goes to zero. One can thus see how the language in which we were historically expressing things was already very close to the large deviation formalism with the variance of the density field at some scale R as the *inverse driving parameter*, $\sigma_\delta^2(R) = 1/n$, that I will from now on just call the driving parameter. It reaches the large deviations regime when going to zero instead of $+\infty$. This small introduction now allows me to express everything in the language of LDT.

Starting from Gaussian initial conditions, the linear density field filtered in a top-hat window of radius r satisfies a large deviation principle with its rate function simply given by a quadratic term in the linear density contrast τ

$$\psi_{\rm ini}(\tau) = \sigma_{l,\delta}^2(R)\frac{\tau^2}{2\sigma_{l,\delta}^2(r)}, \tag{2.30}$$

where we recall that $\sigma_{l,\delta}^2(R)$ is our driving parameter and the variance is equal to the linear variance in the limit where it goes to zero. R is the smoothing radius of the equivalent sphere that contains the corresponding, evolved, non-linear density field and implies the existence of a (one-to-one) mapping from the initial linear to the non-linear field that links r and R and that preserves the spherical symmetry. Indeed, it is shown in Valageas (2002), and even more recently in Ivanov et al. (2019), that for the non-linear density field in spherically symmetric configurations, as is the case when it is filtered in a top-hat window, the leading dynamics (the classical or saddle point path in a path integral formulation of the density PDF among all possible evolutions) is given by the spherical collapse. This notably means that, in the case of the non-linear density field filtered in a top-hat window, the mapping from the linear to the non-linear field that contributes to the infimum of the contraction principle given in equation (2.22) is spherical collapse. This mapping allows us to link R and r through $r = R\rho^{1/3}$‡ (mass conservation) where ρ is the non-linear density field and we

*These are simple numbers only in a EdS space-time. In general they are numbers that also depend on the background cosmological parameters. This dependence is very weak for a ΛCDM universe (Bernardeau et al., 2002).

†The case of primordial non-Gaussianities can also straightforwardly be accounted for in our large deviation formalism as well as in SPT. For LDT, it amounts to modifying the rate function of the linear density field and then proceeding as in the Gaussian case as shown by Uhlemann et al. (2018b) and Friedrich et al. (2019), with the latter being expressed in a language equivalent to LDT but without explicitly mentioning it.

‡For 2D filters we have $r = R\rho^{1/2}$. For more details on spherical collapse dynamics see appendix A.3

parametrise the spherical collapse with

$$\zeta(\bar{\tau}) = \rho = \left(1 - \frac{\bar{\tau}}{\nu}\right)^{-\nu}, \tag{2.31}$$

where $\bar{\tau}$ is the linear density contrast obtained through the most probable mapping between the linear and late-time density fields and ν is chosen so as to reproduce the value of the tree-order density skewness in SPT. For 3D top-hat-filters applied on the field we have $\nu_{3D} = 21/13$ and for 2D $\nu_{2D} = 1.4$. More details on the explicit resolution of the spherical collapse and the parametrisation I use are given in the appendix A.3. This thus allows us to explicitly write down the rate function of the non-linear density field seen as random variable satisfying a large deviation principle:

$$\psi_\rho(\rho) = \frac{\sigma^2_{l,\delta}(R)}{2} \frac{\bar{\tau}^2(\rho)}{\sigma^2_{l,\delta}(R\rho^{1/2})} = \frac{\sigma^2_{l,\delta}(R)}{2} \frac{\left(\zeta^{-1}(\rho)\right)^2}{\sigma^2_{l,\delta}(R\rho^{1/2})}. \tag{2.32}$$

Similarly to the case of "one cell" of the density field, we can also establish a large deviation principle for the random vectors consisting of the non-linear density field filtered in multiple concentric spheres of radii R_i. Since the variance of all those cells is the same in the zero limit, we can indeed identify a single driving parameter, here $\sigma^2_{R_1}$, which is the variance within the smallest sphere. Again starting by expressing the rate function for the linear density contrast and admitting that the most probable mapping between the linear and the non-linear fields is spherical collapse, we obtain the rate function of the late-time density field in concentric spheres of radii R_i

$$\psi_\rho(\{\rho_i\}) = \frac{\sigma^2_{l,\delta}(R_1)}{2} \sum_{k,j} \Xi_{kj}(\{\tau_i\}) \bar{\tau}_k(\rho_k) \bar{\tau}_j(\rho_j), \tag{2.33}$$

where $\Xi_{kj}(\{\tau_i\})$ is the inverse of the linear covariance matrix between the linear density field inside the initial disks (before collapse) of radii $r_k = R_k\, \rho_k^{1/3}$, and $\bar{\tau}_k$ are the linear density contrasts obtained through the most probable mapping between the linear and late-time density fields.

Finally, as another straightforward consequence of the contraction principle, the rate function given by equation (2.33) is also the rate function of any monotonic transformation of ρ, such that for the density contrast $\delta = \rho - 1$, we have $\psi_\delta(\delta) = \psi_\rho(\rho(\delta))$.

2.5.2 Extending to values $\sigma^2_\delta \neq 0$

Since our driving parameter is the variance of the density field, the SCGF of the density field is actually the generating function of the reduced cumulants, the S_p parameters in the limit of zero variance. Those are moreover just simple numbers and a (top-hat) smoothing term since those are given through spherical collapse dynamics, which is equivalent to tree-order perturbation theory. Because of the link of the SCGF with those objects from perturbation theory, we are lead to assume that the SCGF might be expandable in power series with coefficients the S_p parameters. As such we assume that the SCGF is differentiable everywhere, which thus translates to the rate function being convex by result of the Gärtner-Ellis theorem. If the rate function is thus assumed to be convex, then the SCGF is

given by its Legendre transform through Varadhan's theorem and we furthermore assume, for consistency, that the ρ_c critical values closest to the origin at which the second derivative (2.32) reaches zero are assumed to be the starting point after which the rate function is linear which then drives a pure exponential decay in the tails of the PDF in the large deviation regime. From the general case of equation (2.33) we thus obtain the SCGF of the density field filtered in multiple concentric spheres as

$$\varphi_\rho(\{\lambda_i\}) = \sum_i \lambda_i \rho_i - \psi_\rho(\{\rho_i\}), \tag{2.34}$$

with the stationary condition – called this way because it expresses the saddle point approximation in our derivation of Varadhan's theorem – written as

$$\lambda_k = \frac{\partial \psi_\rho(\{\rho_i\})}{\partial \rho_k}, \quad \forall k \in \{1, \cdot, N\}. \tag{2.35}$$

Equation (2.34) fully expresses how to compute all the one-point statistics (actually only the SCGF) of the density field smoothed in one or multiple concentric spheres in the limit where the variance goes to zero (which is equivalent to tree-order SPT for the S_p parameters), with the exception of the variance itself which is by definition not given by the formalism. However, contrary to for example usage of large deviation theory in the context of statistical physics, where we might always be interested in the macro-behaviour, some sort of thermodynamic limit where the number of particles tends to infinity, the cosmological application is however interested in the regime where the the variance takes larger and larger values throughout the evolution of the gravitational instability (*i.e.* the non-linear regime). The main "trick" of large deviation theory in the cosmological context is thus rather to find a motivated prescription to extend results strictly valid in the large deviation limit to non-zero values of the variance of the field. In the one-cell case the solution can be seen as rather simple, as we simply get rid of the limit in the definition (2.7), and take the driving parameter, the variance of the non-linear density, as a free parameter of the theory – for example obtained through high-order perturbative computations or measured in numerical simulations or even real data – and thus obtain the one-cell CGF as

$$\phi_\rho(\lambda) = \frac{1}{\sigma^2_{nl,\delta}} \varphi_\rho\left(\sigma^2_{nl,\delta}\lambda\right). \tag{2.36}$$

This construction thus gives the non-linear variance and the S_p parameters of the spherical collapse dynamics, that is equivalently the CGF/PDF of tree-order Eulerian perturbation theory re-scaled by the non-linear variance. Note that we do not know how to compute the PDF in SPT other than by computing cumulants one by one and performing approximations like the Edgeworth approximation (see appendix A.1.3). In this sense we have already gained something as this approach is thus equivalent to having an infinite Edgeworth expansion for the PDF, and thus it defines a real PDF with physical meaning coming from the spherical collapse dynamics.

For the case of multiple concentric cells, things are unfortunately not so clear. Indeed one could also re-scale the SCGF by the driving parameter but, first there is an arbitrary choice in the choosing of the driving parameter since each variance at each scale behaves the same

way in the zero limit, and then whatever this choice is, the quadratic terms that represent the Gaussian contribution in the power series expansion of the constructed CGF are not exactly accurate, which then becomes the main source of disagreement between this model and measurements that can for example be made in numerical simulations. Indeed remember that spherical collapse dynamics, somewhat linked to tree-order perturbation theory, has a validity regime* limited to quasi-linear scales and redshifts where the PDF is still approached by some sort of Edgeworth expansion around a Gaussian, which makes it important to get the Gaussian contribution right, even before mentioning the accuracy of high-order cumulants. The proposed solution in this case often revolves around fitting the best value of the driving parameter that minimises the residual between a simulated PDF and the theory. In the context of the aperture mass, I propose in section 4.2.2 another solution based on a guess of what the rate function might look like in the non-zero limit of the variance.

2.5.3 Large deviation principle for the log of the density field

From the contraction principle, it is straightforward to obtain the rate function of any monotonic transformation $\mu(\rho)$ of the density field since we thus have $\psi_\mu(\mu) = \psi_\rho(\rho(\mu))$. However, though in the large deviation regime the consideration of one variable or the other is absolutely equivalent, this is not the case in our extension of the formalism to non-zero values of the driving parameter. One question thus arises, is there a better suited variable for which the extension to a non-zero value of the driving parameter is valid for a maximum range of ρ values? Indeed, extending the LDT results to non-zero values of σ^2_μ leads to constant values of the S^μ_p parameters for μ (again up to a smoothing term), which in turns implies non-constant values of the S^ρ_p parameters for the density field. For the sake of clarity, let me rephrase that last point. Extending the results of large deviation theory down to finite values of the variance can be done by many means. One of them is re-scaling the SCGF by the "true" value of the driving parameter, thus in essence keeping the functional form of the non-linear density rate function by neglecting all corrections that might appear but that are sub-exponential in $1/\sigma^2_\delta$. Another way to extend the results of large deviation theory is by noticing that these results can be extended to any function $\mu(\rho)$ of ρ by means of the contraction principle, and again re-scaling by the "true" value of the non-linear variance, but this time for μ. This leads to effectively changing the functional form of the non-linear density rate function outside of the large deviations regime, that is, for finite values of the variance. As such, with the right change of variable, we can incorporate extra non-linear corrections to the PDF of the density field. This idea was introduced in Uhlemann et al. (2016) and has lead to many developments that I will mention in the next subsection. Here I will focus on two aspects that I find important, that a given transformation will allow

*The validity regime of spherical collapse dynamics/tree-order SPT for the S_p parameters is not precisely established other than by numerical experiments (theoretical estimation of SPT loop corrections exist as in Fosalba & Gaztanaga (1998) but are hard to use to probe the precise validity regime) and part of my results in this thesis indeed probe this validity regime by comparing our theoretical PDFs and CGFs to numerical simulations. It should however be noted that it is a specificity of ΛCDM that tree-order SPT is accurate over a somewhat large range of scales and redshifts (up to ~ 10 Mpc/h at $z = 0$). This was shown first in Bernardeau (1994) where for a power-law power spectrum of index $n < -1$, the tree-order S_p parameters were found to be accurate. For our scales of interest and within ΛCDM with typical values of the cosmological parameters, we can expect $n \sim -1.5$.

us to arrive at an analytical expression of the PDF, and that this can lead to meaningful correction to the tree-order SPT result coming from LDT.

One meaningful transformation is the inverse of the spherical collapse, since it is Gaussian and thus preserves the functional form of its rate function even when stepping further away from the large deviations regime. On the other hand, since the natural logarithm is a fairly close function to the inverse spherical collapse, especially for small values of the density, it also makes for a very good mapping and actually, historically, it is because the log was found to work so well that the reasoning about the inverse spherical collapse as a mapping began to take form.

Thus let me define $\mu = \log(\rho)$ such that the SCGF of the μ random variable is given by

$$\varphi_\mu(\lambda) = \lambda\mu - \frac{\sigma^2_{l,\delta}(R)}{2}\frac{\bar{\tau}^2(\rho(\mu))}{\sigma^2_{l,\delta}(R\rho(\mu)^{1/2})}. \tag{2.37}$$

Again the re-scaling of the SCGF to obtain the CGF is done by the non-linear variance but this time of the μ field, which leads for the PDF to constant S^μ_p parameters with respect to scale and redshift and up to a smoothing term. Also note that there is this time no justification coming from perturbation theory to justify the induced functional of the S^ρ_p parameters of the non-linear density field, only the reasoning based on the inverse spherical collapse and of course comparison to numerical simulations. The only thing that is sure is that this is correct in the large deviations regime. Nevertheless, let me give a few interesting results that one obtains from expressing and then extending the large deviation principle at play for the log of the density field.

First let me give explicitly the link between moments and cumulants (connected part) of the moments as

$$\langle\rho^n\rangle_c = \sum_{k=1}^{n}(-1)^{k-1}(k-1)!B_{n,k}\left(\{\langle\rho^m\rangle\}_{m=1,\cdots,n}\right), \tag{2.38}$$

where $B_{n,k}$ are the partial Bell polynomials. Then let me take into account that if $\mu = \log(\rho)$, then $\langle\rho^m\rangle = \langle\exp(m\mu)\rangle = \exp(\phi_\mu(m))$. Then normalising the previous equation to enforce that $\langle\rho\rangle = 1$ we get

$$\langle\rho^n\rangle_c = \frac{\sum_{k=1}^{n}(-1)^{k-1}(k-1)!B_{n,k}\left(\{\exp\phi_\mu(m)\}_{m=1,\cdots,n}\right)}{[\exp\phi_\mu(1)]^n}. \tag{2.39}$$

Finally expanding the CGF of μ in a power series we obtain

$$S^\rho_n = \frac{\exp\left[(n-2)\sum_{l=2}^{\infty}S^\mu_l\sigma_\mu^{2(l-1)}\frac{1}{l!}\right]\sum_{k=1}^{n}(-1)^{k-1}(k-1)!B_{n,k}\left(\left\{\exp\left(\sum_{l=2}^{\infty}S^\mu_l\sigma_\mu^{2(l-1)}\frac{m^l}{l!}\right)\right\}_{m=1,\cdots,n}\right)}{\left[\exp\left(\sum_{l=2}^{\infty}S^\mu_l\sigma_\mu^{2(l-1)}\frac{2^l}{l!}\right)-\exp\left(2\sum_{l=2}^{\infty}S^\mu_l\sigma_\mu^{2(l-1)}\frac{1}{l!}\right)\right]^{n-1}}. \tag{2.40}$$

Most notably, we obtain (Uhlemann et al., 2016)

$$\begin{aligned}\sigma_\rho^2 &= \sigma_\mu^2 + \left(S_3[\mu] + \frac{1}{2}\right)\sigma_\mu^4 + \mathcal{O}\left(\sigma_\mu^6\right),\\ S_3[\rho] &= S_3[\mu] + 3 + \sigma_\mu^2\left(\frac{3}{2}S_4[\mu] + 2S_3[\mu] - 2\left(S_3[\mu]\right)^2 + 1\right) + \mathcal{O}\left(\sigma_\mu^4\right),\\ S_4[\rho] &= 16 + 12S_3[\mu] + S_4[\mu] + \mathcal{O}\left(\sigma_\mu^2\right),\end{aligned} \tag{2.41}$$

which can be finally combined to reach

$$S_3[\rho] = S_3^{\text{tree}}[\rho] + \sigma_\rho^2\left(\frac{3}{2}S_4^{\text{tree}}[\rho] - 4S_3^{\text{tree}}[\rho] - 2\left(S_3^{\text{tree}}[\rho]\right)^2 + 7\right). \tag{2.42}$$

This last result explicitly shows how the large deviation principle applied to different, and carefully chosen, monotonic functions of the density field can yield corrections to the tree-order expression of the reduced skewness of the density field. Now unfortunately these exact additional terms cannot be compared to exact loop corrections in SPT (because they are not yet fully computed) but they nevertheless come from an interesting non-perturbative approach inspired by large deviation results. This is all the more interesting noticing that the CGF is itself an observable just like the PDF since it is defined as the ensemble average of the exponential of the field times λ which means that those corrections can become meaningful for real observations and thus tested apart from their influence on the PDF. Comparisons of this result with numerical simulations are, as mentioned previously, found in Uhlemann et al. (2016).

The last point of this subsection on the log of the density field concerns the actual computation of the PDF rather than the CGF. Note that this computation is mandatory only because we cosmologists as a field are not used to dealing with CGFs rather than PDFs. However, note that CGFs formally offer no formal advantages or drawbacks compared to the PDF, not in terms of information content nor measurement constraints and systematics, and we focus on the PDF only because ideas are more easily expressed and transmitted in terms of PDFs rather than CGFs – think for example of intervals of probability or cuts in the tails of the PDF which we feel we control more when dealing with the PDF.

Formally, the PDF is expressed as the inverse Laplace transform of the CGF, which thus need to be analytically continued in the complex plane

$$\mathcal{P}(\rho) = \int_{-i\infty}^{+i\infty}\frac{\mathrm{d}\lambda}{2\pi i}\exp\left[-\lambda\rho + \phi_\rho(\lambda)\right]. \tag{2.43}$$

Several methods to perform the continuation and then numerically carry out this type of integral will be given in section 3.5 and I instead here want to focus on what can be done analytically and what changes when working with μ rather than ρ. Indeed, taking advantage of the fact that the non-linear variance, be it for μ or ρ, is relatively small we might, as in section 3.5.1 where I derive it, express the PDF as a saddle point approximation which leads to

$$\mathcal{P}(\rho) = \sqrt{\frac{\psi_\rho''[\rho]}{2\sigma_{nl,\rho}^2\pi}}\exp\left(-\psi_\rho[\rho]/\sigma_{nl,\rho}^2\right). \tag{2.44}$$

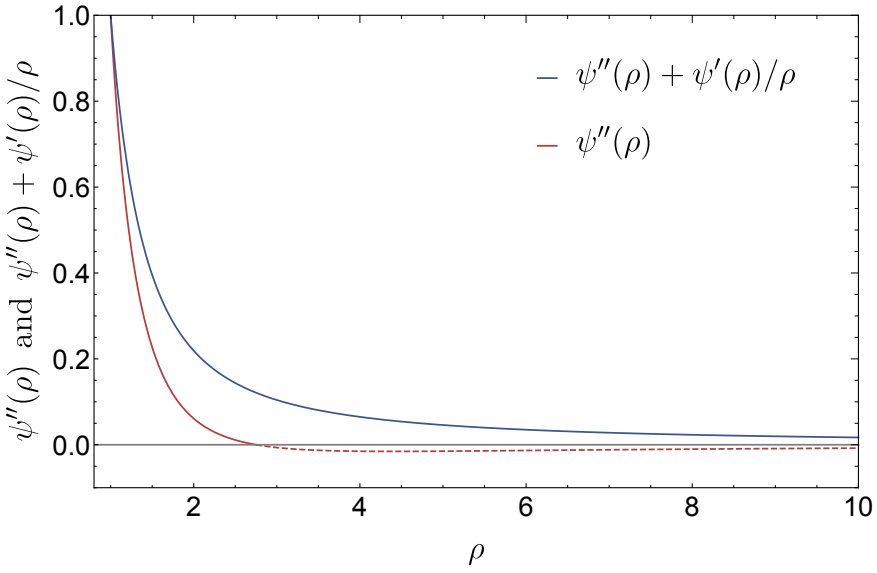


Figure 2.3: Comparison of the argument of the square-root in the PDF of the density field whether the expression is obtained directly from a large deviation principle of the density field ρ and a saddle point approximation (red line) or from a large deviation principle of the log of the density field μ (blue line). We observe that indeed in the critical point is pushed further away when using the log mapping rather than the density field directly which then allows us to obtain a good analytical expression for the PDF of the density field. The dashed line indicates that the rate function after the critical point is considered linear and thus negative values are deemed nonphysical and not taken into account in the formalism.

Unfortunately, though accurate with respect to the numerical integration for $\rho < \rho_c$, the critical point at which $\psi''(\rho_c) = 0$, the expression above is not valid (because the real square-root cannot admit a negative argument) beyond or even near the critical point. Thus though the critical point is not in itself an issue since it does not correspond to any singular behaviour in the CGF, it is just a linear branch in the rate function, it prevents us from computing an accurate analytical estimate of the PDF if, and it is, the critical point is reached for rather low values of the density field. I illustrate this in figure 2.3. On the other hand, performing the exact same saddle point approximation for the variable μ and obtaining the same result (only changing ρ in μ every time it appears) leads to a different PDF for the density field itself. Indeed, noticing that $\mathcal{P}_\rho(\rho)\mathrm{d}\rho = \mathcal{P}_\mu(\log(\rho))\mathrm{d}\rho/\rho$, we obtain that

$$\mathcal{P}(\rho) = \sqrt{\frac{\psi''_\rho[\rho] + \psi'_\rho[\rho]/\rho}{2\sigma_\mu^2\pi}} \exp\left(-\psi_\rho[\rho]/\sigma_\mu^2\right). \tag{2.45}$$

What changes is that thanks to the μ mapping, the range of ρ values for which $\psi''_\mu(\mu) > 0$ is then considerably increased, which translates into a much bigger range of ρ values in the PDF for which the argument of the square root remains positive. I again illustrate this in figure 2.3.

As a consequence and to conclude this subsection, the equivalent statement that μ and ρ have the same large deviation behaviour (same rate functions expressing ρ as $\rho(\mu)$ and vice versa) leads to quite different statements once we try and extend the large deviations results near but outside the large deviations regime for finite values of the variance. Indeed we saw

that this leads to different cumulants for the density field, for which I tried to justify how they are more meaningful, and thus different PDFs. Moreover, because the critical behaviour of the rate function is changed through different mapping, this is also helpful for obtaining an analytical approximation of the PDF by performing yet another saddle point technique.

2.5.4 Results and the cosmological large deviation principle in the literature

To conclude this rather introductory chapter on the application of the large deviation principle in cosmology, I compare the results of the numerical integration of equation (2.43) and the PDF (2.45) obtained by a saddle point approximation and re-normalised to enforce $\int \mathcal{P}(\rho)\mathrm{d}\rho = 1$ and $\langle\rho\rangle = 1$, which is not exactly preserved when performing the saddle point approximation. Additionally, and to also illustrate the validity of the formalism, I compare the computed PDFs with that measured in the state-of-the-art simulation Horizon Run 4 (Kim et al., 2015). I choose a sphere radius of 15 Mpc/h and 3 different redshifts, $z = 0, 0.7$ and $z = 4$. One can then observe in figure 2.4 that in both cases, with LDT applied to both ρ (dashed lines) and μ (solid lines), the agreement between the induced theoretical density PDF is in very good agreement with the measurements in the simulation, even in the low probability density tails. Moreover, and as expected from our formalism, the PDF obtained from the μ mapping appears to be even more successful at describing the rare events tails of the density PDF. Nevertheless, remember that the success of this approach illustrates in some sense a particularity of the ΛCDM universe for which tree-order SPT is accurate down to quite small scales ($\sim$ 10 Mpc/h), which thus also implies that loop corrections (I am talking only at the level of reduced cumulants and not the power spectrum for example) are rather small. This was already hinted all the way back in 1998 in Fosalba & Gaztanaga (1998).

Finally, let me note that the ideas I have introduced in this chapter have led to many different results all building on these basic tools. For example, the computation of the joint statistics of concentric spheres, for which I have presented the theory but no illustration, was also performed in Uhlemann et al. (2016) still making use of the logarithm of the density field. The inclusion of non-Gaussian initial conditions is, as already mentioned, performed by modifying the initial non-Gaussian rate function by adding a small but non-zero f_{NL} term (only local non-Gaussianities contribute to one-point statistics) and still using spherical collapse to map the linear to non-linear density field (Uhlemann et al., 2019). The spherical collapse model also provides a means to link the linear and non-linear velocity field, and thus it is also possible to study the joint PDF of the density and velocity fields as in Uhlemann et al. (2017b) in the LDT framework. Given a bias expansion model to link the matter density field to halos or galaxies, we can then from our PDF of the density field express via a change of variable the halo or galaxy PDF with quite good accuracy (Uhlemann et al., 2018a), and the same idea can be exploited for 21cm intensity mapping fitting in hydro-dynamical simulations a bias function between the density field and neutral hydrogen (Leicht et al., 2019). One very interesting idea that also emerged and that is, from my point of view, one of the most interesting yet since the PDF of the density itself, is the proof that that the joint PDF of distant spheres can be computed directly in terms of PDFs of concentric spheres and a bias function which is expressed at leading order in terms of the cumulant generating function of a single cell. Indeed the works in Codis et al. (2016) and

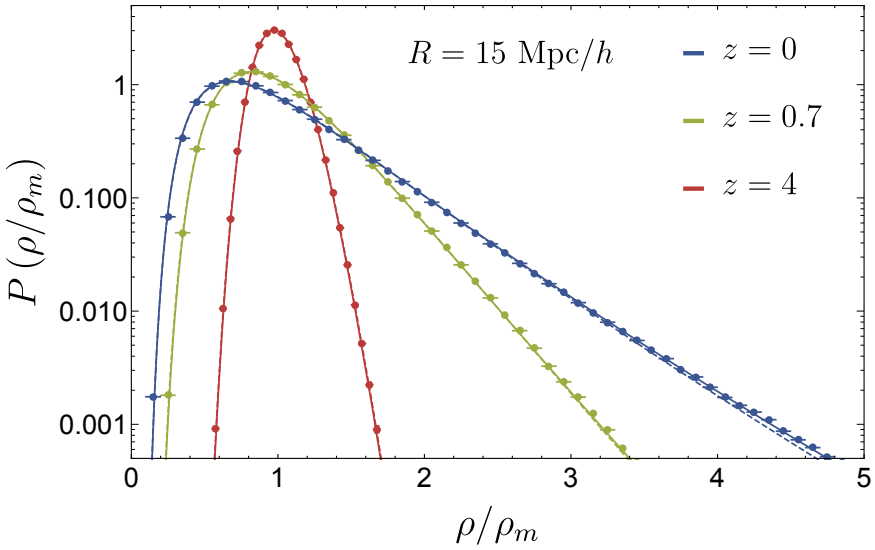


Figure 2.4: PDF of the density field filtered in a top-hat window of radius 15 Mpc$/h$ at different redshifts. The solid lines correspond to the PDF obtained from the application of the large deviation principle to the log of the density field and the dashed lines directly to the density itself. The point with error bars are measured in the Horizon Run 4 simulation, vertical bars display the error on the mean among 8 sub-cubes to probe the cosmic variance and the horizontal bars show the binning in which the PDF (normalised histogram) is measured.

Uhlemann et al. (2017a) (respectively without and with the μ mapping) exploit the idea that for 3 concentric spheres of radii $R_1 << R_2 \lesssim R_3$ whose joint PDF we know, we then also know the joint statistics of the density in the sphere R_1 and the spherical shell $R_3 - R_2$, and thus, following isotropy in the universe, which thus means that the additional sphere can be anywhere, this is equivalent to knowing the joint PDF between the sphere R_1 and a sphere containing the matter in $R_3 - R_2$ separated by a distance R_2. This idea, which yields results valid strictly speaking only at large separation and at leading order, is key to establishing a theoretical model for the covariance matrix of the density PDF, which has a key role in the estimation of the cosmological information present in the density PDF.

In the next chapters, I will extend the LDT formalism to weak-lensing observables. I will first deal with the PDF of the convergence field in chapter 3, show how one obtains a meaningful and accurate PDF in this formalism, treat the addition of standard observational systematics on top of our formalism and study the cosmological information content of the convergence PDF as an observable modelled by our formalism. In chapter 4, I will move from the convergence to the more observable aperture mass and introduce how one moves beyond top-hat smoothing in the LDT formalism. This last chapter also contains the addition of the BNT transform to the lensing formalism, which is independent of our large deviations formalism.

3

A first step towards realistic theoretical lensing distribution functions: the convergence PDF

Contents

3.1. Introduction and non-Gaussian statistics in the literature

As mentioned in chapter 1, the statistics of weak-lensing fields provide a powerful tool for precision cosmology while, by essence, the weak nature of the effect renders viable its exploitation only in large area surveys observing many millions to billions of galaxies. As such this motivated the build up of new generation large galaxy surveys such as the Dark Energy Survey (DES, DES Collaboration, 2018), the Kilo-Degree Survey (KiDS, Heymans et al., 2021) and the Hyper Suprime Cam (HSC, Hikage et al., 2019), and in the near future even larger and more powerful surveys such as Euclid (Laureijs et al., 2011; Amendola et al., 2018) and the Rubin Observatory (Legacy Survey of Space and Time – LSST) (Ivezić et al., 2019). This huge quantity of data thus renders mandatory the use of accurate statistical probes able to infer the matter distribution between us and the sources and it is thus the time to survey the literature on the most commonly studied statistics, their constraints on cosmological parameters and hence our motivations for developing models of the lensing PDF over the course of this thesis.

A large majority of the literature focuses on the information contained in power spectra or equivalently their real-space counterparts the two-point correlation functions, and as far as cosmic shear experiments are concerned the master-tool is given by a combination of three 2-point functions – the so-called *3×2 points analysis*:

Galaxy Clustering: The two-point function between galaxy positions in redshift bins i and j describing the excess (over random) number of galaxy pairs separated by an angular distance θ. It is common to only use the auto-correlations ($i = j$) in the analysis.

Galaxy–Galaxy Lensing: The two-point function between galaxy positions and galaxy tangential shear in redshift bins i and j which describes the over-density of mass around galaxy positions since the matter correlated with the source galaxy alters the path of the light emitted from it.

Cosmic Shear: The two-point correlation between galaxy shear in redshift bins i and j described by two functions which are the sum and difference of the products of the tangential

and cross-components of the shear.

A very recent and state-of-the-art analysis of this type can be found in the results of the third year of the exploitation of the Dark Energy Survey (DES Collaboration et al., 2021). Most notably, cosmic shear experiments probe the clustering of matter and are thus mostly sensitive to the fraction of matter in the energy budget of the universe Ω_m, the linear root-mean-square of matter fluctuations smoothed in a top-hat of comoving radius 8 Mpc/h σ_8, which can be seen in the overall normalisation of the linear matter power spectrum, and more weakly to the dark energy equation of state and the sum of neutrino masses. Uniquely using its 3×2 point analysis, the DES collaboration for example obtained $\Omega_m = 0.352^{+0.035}_{-0.041}$ and $\sigma_8 = 0.719^{+0.037}_{-0.044}$. Note that these constraints improve by a factor of 10 when combining the DES data with external data such as the Big-Bang nucleosynthesis (BBN), cosmic microwave background (CMB) temperature and polarisation measurements from the Planck mission (Planck Collaboration et al., 2020a) or the baryonic acoustic oscillations (BAO) peak.

However, we know that the power spectra contain only complete statistical information for Gaussian random fields, a prescription valid with extremely good accuracy to describe primordial metric perturbations visible in the CMB as shown by Planck Collaboration et al. (2020a), but not for the subsequent *non-linear* time-evolution of density fluctuations by means of the gravitational instability which then develop significant non-Gaussianities – in particular for small scales and late times. In this non-linear regime of structure formation, we observe both an increase in power in the matter power spectrum measurements (relative to linear evolution) and a generation of distinct non-Gaussianities in the late-time density field as already seen in the PDF of the matter density field in the previous chapter 2 (and particularly in figure 2.4). This, by projection, also implies strong non-Gaussian features of the weak-lensing fields which should then also carry non-negligible cosmological information complementary to that contained in the power spectra. For example, Zorrilla Matilla et al. (2020) showed with a series of Deep Neural Networks that a significant amount of the cosmological information in the convergence field lies in the extreme rare events, that is in the tails of its PDF.

Many different tools have been put forward over the course of the last decades to study the non-Gaussian aspects of the matter and lensing fields, the most obvious being higher order correlation functions such as the 3-point correlation function or its Fourier counterpart the bi-spectrum, whose definition is analogous to the power spectrum but with three wave numbers instead of two. There are many works that could be cited as an example of the shear bispectrum but I find more interesting to mention Halder et al. (2021) which defines and models a three-point function analogous to the bi-spectrum but more easily measured in data. Alternatively, other popular probes of non-Gaussianity include peak counts (Kacprzak et al., 2016; Shan et al., 2018) where one measures the number of high-valued peaks above a certain threshold ν (or equivalently above a signal to noise ratio) as a function of ν and which thus depend on the values of the field and its spatial derivatives for the peak definition. Minkowski functionals (Marques et al., 2019) concern the area, length of the boundary (sort or perimeter), and genus (sort of the curvature) of the excursion sets formed by the field above a threshold ν. For the case of cosmological constraints coming from lensing peak counts and Minkowski functionals, a recent analysis can be found in Zürcher et al. (2021) where it is shown that the Figure-of-Merit in the $\Omega_m - \sigma_8$ plane increases by a factor

of 5 when adding those non-Gaussian statistics to the standard angular power spectrum. Note however that these quantitative analyses assume that non-Gaussian quantities can be predicted with great precision at the chosen filtering scales, which is not all a given especially when using un-tested – because of the absence of a good theoretical model – predictions from numerical simulations. Martinet et al. (2020) investigated the constraining power of weak lensing aperture mass statistics, such as peaks, voids and the probability distribution function (PDF), while Zürcher et al. (2021) added Minkowski functionnals, and also confirmed that the non-Gaussian statistics could significantly enhance constraints in the $\Omega_\mathrm{m} - \sigma_8$ plane. Finally, and this time on real data, a very recent study by Harnois-Déraps et al. (2021) found one of the tightest constraints on $S_8 = \sigma_8\sqrt{\Omega_\mathrm{m}/0.3}$ in DES year-1 data based on a simulation-based peak count and correlation function combined analysis.

As for probability distribution functions, though less popular, they still receive some attention (see for example Friedrich et al. (2018); Gruen et al. (2018)) and offer the possibility of being much more easily applied to real data. They have also been shown to offer additional constraints compared to the power spectrum on cosmological parameters as is shown for example in Uhlemann et al. (2019) for the matter density PDF and in Boyle et al. (2021) for the convergence PDF, with the latter being one project I worked on during my thesis and that I thus describe in this chapter. In particular, Friedrich et al. (2018) used the cumulant generating function to construct the one-point PDF of galaxy densities in cones and the weak-lensing convergence profile around lines-of-sight with given galaxy density. This density-split statistics from counts and lensing in cells can yield cosmological constraints competitive with the two-point function measurements (see figure 10 in Gruen et al., 2018), and potentially, the lensing profile around special density environments like voids, as suggested for example by Krause et al. (2013), can also be used to test theories of gravity (Baker et al., 2018).

A Fisher analysis based on fast simulations in Patton et al. (2017) demonstrated that the weak-lensing convergence PDF provides information complementary to the cosmic shear two-point correlation and note that this complementarity between two-point and non-Gaussian statistics is even more relevant in the presence of systematics such as shape noise: They showed that although the degeneracy direction between Ω_m and σ_8 was very similar to the cosmic shear two-point correlation function, adding information from the PDF improved constraints by about a factor of two compared to two-point statistics alone. Additionally, the presence of measurement systematics was shown to tilt the degeneracy directions between the PDF and the power spectrum. This is in line with reports of an increase in the lensing figure of merit through higher order convergence moments in Vicinanza et al. (2018) and an improvement by a factor of two when adding moments over shear power spectrum tomography alone (Petri et al., 2016b). The same conclusions were reached by Petri et al. (2013) demonstrating the additional information gained on cosmological parameters (in particular the equation of state of dark energy, amplitude of fluctuations and total matter fraction) by including higher order moments compared to an analysis of the sole power spectrum of weak-lensing data. In addition, numerical simulations suggest that the lensing convergence PDF contains signatures of massive neutrinos, with the PDF alone providing a 20% stronger constraint on M_ν than the power spectrum while combining the two measurements can improve the constraint by 35% over that of the power spectrum alone (Liu & Madhavacheril, 2019). Including non-Gaussian information is also key to breaking degeneracies between

massive neutrinos and modified gravity $f(R)$ models in the weak lensing power spectrum (Peel et al., 2018).

I have already mentioned several times that it is hard to derive accurate theoretical probes of higher-order (non-Gaussian) statistics, and even more so where the cosmological dependence can be accurately derived. As such, the extreme majority of the mentioned and existing analysis relies on numerical simulations for which, though their necessity and quality are indisputable, some hardly-evaluated uncertainties remain, as discussed in section 1.8.

While empirically, the one-point distributions of matter and tracer densities are found to be close to log-normal (Coles & Jones, 1991; Kayo et al., 2001; Bel et al., 2016; Hurtado-Gil et al., 2017; Repp & Szapudi, 2018), similarly to the weak-lensing distribution (Taruya et al., 2002; Hilbert et al., 2011; Clerkin et al., 2017), log-normal models are fundamentally limited in jointly modelling the two (Xavier et al., 2016). Even at the level of the one-point distribution, the log-normal model does not capture the detailed behaviours of the field evolution and several tentative models have been considered to circumvent this issue with skewed log-normal (Colombi, 1994), generalised normal distributions (Shin et al., 2017) or double exponential cutoffs (Klypin et al., 2018) for instance. However, to extract cosmological information a more physically principled model paramaterising the full cosmology dependence of the PDF is required and presents many advantages. It allows for easier intuitive understanding of the cosmological dependence of the statistic, removes the need for analysis with expensive simulations, and greatly simplifies the performance of cosmological survey forecasts. In this direction, the convergence PDF was for example based on the halo model in Thiele et al. (2020).

More physically motivated approaches have also been developed over the course of the last decades. In the quasi-linear regime, perturbation theory can be used for the computation of cumulants and probability distribution functions of the density field via Gram–Charlier A or Gamma expansions, the latter allowing for a strictly positive definite PDF (Gaztañaga et al., 2000). Recently, non-perturbative effects on the dark matter PDF have been quantified analytically in one dimension (Pajer & van der Woude, 2018) and estimated in 3D using a path-integral approach based on perturbation theory and a renormalisation procedure for small-scale physics in Ivanov et al. (2019). Following another direction and in order to get the full hierarchy of cumulants and a meaningful PDF, Bernardeau (1992); Valageas (2002); Bernardeau & Reimberg (2016) suggested the use of the spherical collapse model later re-interpreted in the context of large-deviation theory. In this approach, already presented in the previous chapter 2, the matter PDF is predicted with very good accuracy in the mildly non-linear regime (Bernardeau et al., 2014a; Uhlemann et al., 2018c, 2016, and figure 2.4).

As such, based on **Barthelemy** et al. (2020a), **Barthelemy** et al. (2020b), **Barthelemy** et al. (2021b) and Boyle et al. (2021), and after having introduced the basics of cosmic shear experiments, how the measured quantities depend on the clustering of the matter density field, and how one could model this clustering through the use of the spherical collapse and the large deviation principle in chapters 1 and 2, this chapter now combines everything to present a model for the convergence PDF fully taking into account the geometry of the light-cone and with comparison to numerical simulations. I will also present the first comprehensive forecast for the lensing convergence PDF simultaneously varying all ΛCDM parameters, as well as individually considering the extensions of a free total neutrino mass M_ν and dark energy equation of state w_0. This model is also useful for a more accurate

derivation of the density split statistics presented in Friedrich et al. (2018).

I will start with the introduction of the projection formula that links the projected cumulants to the cumulants of the matter field at different scales and redshifts, discuss the treatment of the driving parameter in this new configuration (is it possible to keep only one?) and then present some comparisons between the hereby modelled CGF and the one extracted from the Takahashi simulation. The next parts will discuss the different strategies one could adopt to perform the inverse Laplace transform of the CGF to obtain the PDF and once again compare the results to the Takahasi simulation. I will discuss how the model can be extended to account for shape noise, additional lensing corrections and massive neutrinos and I will dedicate a section to the PDF dependence on cosmological parameters and the Fisher forecasts one could thus expect for a next generation – Euclid-like – survey. Finally, I will conclude this chapter with the introduction of the PDF of the BNT convergence, which is proven to be fully accurate over the whole range of accessible probability densities.

3.2. The projection formula: from the 3D density field to the weak-lensing cumulants

From equation (1.58), the convergence is seen as coming from the superposition of (supposedly) independent layers of the 3D cosmic matter field. Then making use of the Limber approximation which only considers wave-modes orthogonal to the line of sight, the then-obtained filtered convergence field is seen as a projection of independent thin slices of matter with evolving size along the line of sight as is sketched in figure 1.7. As a consequence the computation of the one-point PDF would involve an infinite number of convolution products between the individual PDFs of the 2D matter field at the relevant scales and redshifts. Alternatively, one can more conveniently consider the cumulants of the lensing field (equivalently the cumulant generating function) which is now simply an addition of the cumulants of the superposed layers. This is the path I choose to follow, the result being explicitly shown in equation (3.3) below. Given the importance of this result and though it could be seen as a strict application of the Limber approximation, let me quickly re-derive it here.

Following Bernardeau & Valageas (2000), let me start by writing explicitly the p-point correlation functions of a projection of density field with kernel F and the subscript c denotes the connected part of the ensemble average, *i.e.* the cumulants

$$\langle \delta_{\rm proj}(\boldsymbol{\theta}_1) \, ... \, \delta_{\rm proj}(\boldsymbol{\theta}_p) \rangle_c = \int_0^{\chi_s} \prod_{i=1}^{p} \mathrm{d}\chi_i \, F(\chi_i) \times \langle \delta(\chi_1, \chi_1\boldsymbol{\theta}_1) \, ... \, \delta(\chi_p, \chi_p\boldsymbol{\theta}_p) \rangle_c. \tag{3.1}$$

This quantity can be computed by making the change of variable $\chi_i = \chi_1 + r_i$. Since the correlation length (beyond which the p-point correlation functions are negligible) is much smaller than the Hubble scale $c/H(z)$, only values of r_i which obey $|r_i| \ll c/H(z)$ contribute to the integral over r_i. Thus the integral boundaries over r_i can be pushed to infinity or to any value greater or equal to the values for which the correlation is negligible. If this radial correlation length is small enough and if the F kernel is slowly varying I obtain

$$\langle \delta_{\rm proj}(\boldsymbol{\theta}_1) \, ... \, \delta_{\rm proj}(\boldsymbol{\theta}_p) \rangle_c = \int_0^{\chi_s} F(\chi_1)^p \, \mathrm{d}\chi_1 \times \int_{-\infty}^{\infty} \prod_{i=2}^{p} \mathrm{d}r_i \times \langle \delta(\chi_1, \chi_1\boldsymbol{\theta}_1) \, ... \, \delta(\chi_1, \chi_p\boldsymbol{\theta}_p) \rangle_c. \tag{3.2}$$

Now taking filtering effects into account, the integrals over r_i will only take the differential elements as integrand and we are finally led to

$$\langle \delta^p_{\mathrm{proj},\theta} \rangle_c = \int_0^{\chi_s} \mathrm{d}\chi \, F^p(\chi) \, \langle \delta^p_{\chi\theta,\mathrm{cyl}} \rangle_c \, L^{p-1}, \tag{3.3}$$

where $\langle \delta^p_{\chi\theta,\mathrm{cyl}} \rangle_c$ are the cumulants of the 3D density contrast filtered in a cylinder of transverse size $\chi\theta$ and (infinite) length L. Note that the length L of the cylinders is only a dimension parameter that will always cancel out when explicitly writing the cylindrical cumulants. Moreover, since the cylindrical collapse we consider is actually a two-dimensional collapse (valid when the length of the cylinder tends to infinity), it is easier get rid of the L and consider that equation (3.3) states that the projected cumulants are a weighted integral of the cumulants of infinitely thin slices of the underlying 3D matter density field.

From there, we easily obtain the projected cumulant generating function

$$\phi_{\mathrm{proj},\theta}(y) = \int_0^{z_s} \frac{\mathrm{d}z}{H(z)} \, \phi_{\mathrm{cyl}}(F(z)y, z). \tag{3.4}$$

3.3. Convergence Cumulant Generating Functions with large deviation theory

3.3.1 The convergence does not satisfy a large deviation principle

A sum of variables satisfying a large deviation principle does not necessarily satisfy this principle itself. For the case of the convergence and later the aperture mass, this in particular means that it might not be possible to formally write down some reduced cumulant generating function in the limit where some driving parameter goes to zero and we would thus have to directly compute the cumulants (as opposed to their reduced counterpart) of the density field to obtain the cumulants of the convergence and aperture-mass field.

To be more explicit, I would like to offer some intuition on why the convergence may not satisfy a large deviation principle*. Let me start by re-stating that for a random variable ρ satisfying a large deviation principle and whose SCGF is differentiable, the following equations hold

$$\begin{cases} \varphi(\lambda) = \lambda\rho - \psi(\rho), \\ \rho = \dfrac{\partial}{\partial\lambda}\varphi(\lambda), \\ \lambda = \dfrac{\partial}{\partial\rho}\psi(\rho) \end{cases} \tag{3.5}$$

where the rate function ψ ceases to be strictly convex at $\rho = \rho_c$ and thus $\lambda_c = \psi'(\rho_c)$ and $\psi''(\rho_c) = 0$. Then expanding the 3rd equation of (3.5) around (λ_c, ρ_c) and integrating for

*The use of the conditional comes from a lack of confidence from my part that this is strictly correct from the mathematical point of view. Indeed what I demonstrate is not that it is impossible to write a rate function for the convergence but rather that the critical behaviour of the convergence if a rate function existed and the SCGF was differentiable is not recovered. This question will be addressed more rigorously somewhere else.

$\varphi(\lambda)$ I obtain

$$\varphi(\lambda) = \varphi(\lambda_c) + (\lambda - \lambda_c)\rho_c + \frac{2}{3}\sqrt{\frac{2}{\psi'''(\rho_c)}}(\lambda - \lambda_c)^{3/2} + \cdots \tag{3.6}$$

where the $(\lambda - \lambda_c)^{3/2}$ term illustrates the general behaviour of the (scaled or not) generating function around a critical point. This computation is rather general and for a unimodal PDF gives direct information on the probability decay of rare events as a function of the rate function. Bernardeau et al. (2014a) for example showed that this leads to the PDF exhibiting an exponential cut-off in $P_\rho(\rho) \sim \exp(-\lambda_c\rho)/(\rho - \rho_c)^{5/2}$.

Now let us take a look at the rare-event-behaviour of the convergence PDF. Looking at the third derivative of the projected generating function and following equation (3.4), it is the projection of the third derivative of the generating function of the matter field along the line of sight with some kernel resulting from the derivatives. At a particular redshift along the line of sight, when λ is such that $\omega(z)\lambda$ approaches a critical point λ_c of the matter field at this scale and redshift, the third derivative of the matter field diverges as $(\lambda - \lambda_c)^{-3/2}$ so that the third derivative of the projected generating function is dominated by large values of the lensing kernel. We thus have

$$\phi^{(3)}_{\rm proj}(\lambda) \sim \int_{-\infty}^{+\infty} \mathrm{d}\chi[(1-\chi^2)\lambda - \lambda_c]^{-3/2}. \tag{3.7}$$

This can be integrated and leads to

$$\phi^{(3)}_{\rm proj}(\lambda) \sim \frac{1}{\lambda - \lambda_c}, \tag{3.8}$$

which then implies that the generating function approaching a critical point behaves as

$$\phi_{\rm proj}(\lambda) \sim (\lambda - \lambda_c)^2 \ln(\lambda - \lambda_c). \tag{3.9}$$

This behaviour obviously leads to a different rare-event tail for the convergence PDF and thus illustrates how the convergence (and similarly the aperture mass) cannot satisfy a large deviation principle if the matter density field does.

3.3.2 Driving parameters along the line of sight

We have seen in the previous chapter 2 that the large deviation formalism allows one to compute the scaled cumulant generating function of the matter density field in the limit where the driving parameter, the variance of the field, goes to zero. Furthermore we know from the previous subsection that the convergence does not satisfy a large deviation principle, so we first need to extend the large deviation results to the non-zero variance case for the matter field before accessing the results for the convergence field. In practice this means that the exterior (but physical) input needed to model the convergence field with large deviation theory is no longer just the non-linear variance at one pivot scale and redshift but the whole non-linear power spectrum across all probed scales and redshifts along the line of sight.

The need for the external specification of the non-linear variance along the line of sight does somewhat weaken the "from first principles" quality that the large-deviation framework

in cosmology usually exhibits, but nonetheless note that the non-linear power spectrum has been the focus of a lot of attention in recent years and is very reliably modelled now, be it with emulator approaches calibrated with N-body simulations (Takahashi et al., 2012; Euclid Collaboration et al., 2019) or with fully perturbative approaches (Taruya et al., 2012).

Another approach could be to still crudely model the non-linear variance with the help of a somewhat pivot scale. However, because we also want to impose the correct variance for the convergence itself, the re-scaling would be so that the projected variance is correct. In this scheme we have

$$\sigma_{\delta,nl}(R,z) = \alpha\,\sigma_{\delta,l}(R,z) \tag{3.10}$$

where nl and l respectively stand for non-linear and linear, while R and z stand for the scale and the redshift. Then imposing the correct projected variance would then implicitly fix α through

$$\phi_{\kappa,\theta,nl}(y) = \left(\frac{\sigma_{\kappa,l}(\theta)}{\sigma_{\kappa,nl}(\theta)}\right)^2 \phi_{\kappa,\theta,l}\left(\left(\frac{\sigma_{\kappa,nl}(\theta)}{\sigma_{\kappa,l}(\theta)}\right)^2 y\right). \tag{3.11}$$

Here $\phi_{\kappa,\theta,l}$ is the convergence cumulant generating function obtained through equation (3.4) and considering the linear variance as driving parameter along the line of sight, that is without re-scaling by the non-linear variance in each redshift slice along the line of sight. This approximation allows us to keep only one external parameter – the non-linear variance of the convergence field – by considering that the scale and redshift dependence of the linear power spectrum is correct up to a normalisation. This is obviously not true but works very well for the convergence field, and even better for the BNT convergence, thanks to the lensing kernel peaking at some redshift thus giving an important weight to only a short range of scales and redshifts. Typically, the convergence cumulants only differ by a few percent using the full non linear power spectrum or this approximation, and this few percent does not affect the convergence PDF sufficiently to produce any noticeable differences for the values of the probability density that we look at (a few orders of magnitude in probability away from the maximum values reached by the PDF). This somewhat simplifies the practical implementation because, though it could be accessed by other means, it allows me to simply measure the non linear variance from the numerical simulation I compare this formalism to and thus only measure differences coming from the validity of both the simulation and the spherical/cylindrical collapse dynamics on higher order cumulants.

3.4. Simulated vs. theoretical CGFs

Now that we are able to compute the full cumulant generating function of a 2D slice of the matter field by means of large deviation theory and thus the full CGF of the convergence field by means of the projection formula (3.4), I now display a comparison of the theoretical result to the result obtained in the Takahashi simulation described in section 1.8.2. For this comparison, because one full sky provides enough statistics and because processing them all would require an un-necessary amount of numerical ressources, I only pick one of the 108 available full-sky realisations of the convergence field at different redshifts. A glimpse of a convergence map at redshift 2 from this realisation was already shown in figure 1.10.

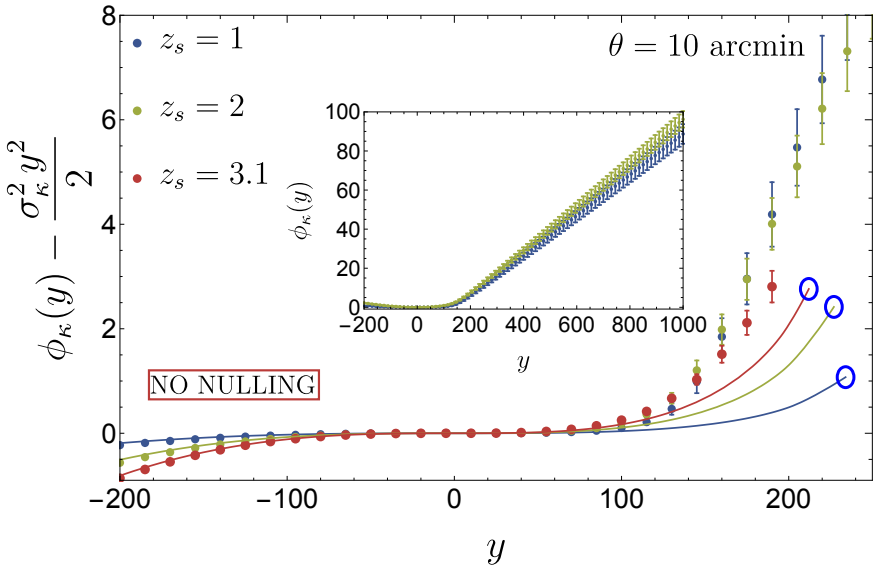


Figure 3.1: CGF of the convergence for different source redshifts and an opening angle of $\theta = 10$ arcmin. Data points were measured on the Takahashi simulation. Blue circles indicate the critical points along the real-axis. The small panel inside shows the linear asymptotic behaviour of the measured CGF. The error bars are obtained from the error on the mean among 8 subparts of the full sky. This figure was published in **Barthelemy** et al. (2020a).

Remember that the (S)CGF is an observable on its own as it is defined by $\phi_\kappa(y) = \log\left(\langle e^{y\kappa}\rangle\right)$ and I thus display in figure 3.1 the convergence CGF for three source redshifts and one opening angle $\theta = 10$ arcmin. I subtract on this plot the Gaussian (quadratic) contribution to the CGF as the mean is imposed by definition to be zero and the variance is chosen as the exterior parameter to match the data variance. Hence, we are only comparing contributions coming from the skewness, kurtosis and higher order cumulants. As already noticed in section 3.3.1, the presence of critical points (non-convexity of the rate function) for the matter field translates into the emergence of a critical value (blue circles on figure 3.1) along the real axis for the projected SCGF/CGF. In principle, those critical points should also have an observational interpretation in observed or simulated data and manifest as a drastic rise of the error bars to a point where the measured signal does not make sense anymore, as was shown in the 3D case in Bernardeau et al. (2014a). This is because above the critical value, ensemble averages of $\exp(y\kappa)$ formally diverge. In practice they are finite in a finite sample but dominated by the rarest events available. More precisely the maximum value of κ available in the sample, $\kappa_{\max}$, gives the CGF a linear asymptotic behaviour of slope $\kappa_{\max}$ and a linear asymptotic for its r.m.s value of slope given by the r.m.s. value of $\kappa_{\max}$ in subsamples. Overall, we observe a rather good agreement between theory and simulated data close to $y = 0$, however the agreement degrades towards the tails when the error on the skewness starts to be visible together with the addition of higher order contributions. This is due to the lensing kernel that mixes all scales and in particular the small ones at the tip of the cone that cannot be well modelled by the spherical collapse dynamics that is here equivalent to tree-order perturbation theory up to the re-scaling by the κ non-linear variance. This might even be clearer when I display the comparison of the theoretical and

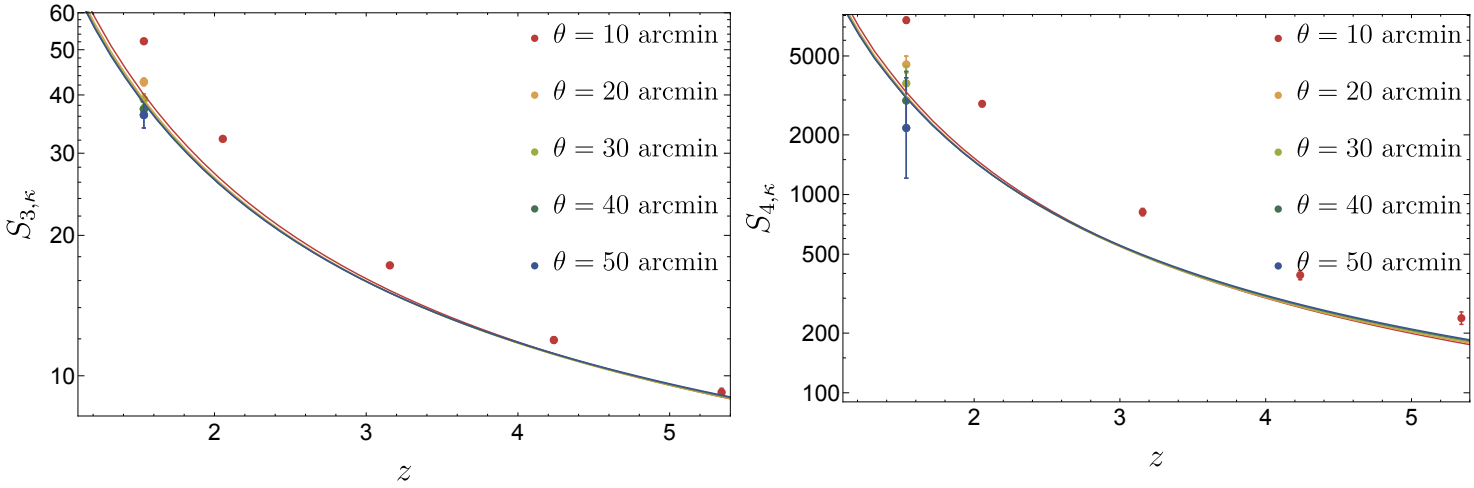


Figure 3.2: High order cumulants of the weak-lensing convergence as a function of redshift and for different opening angles from 10 (blue) to 50 arcmin (red). The solid lines display the predictions extracted from the CGF while the measurements on the simulated (Takahashi's) sky are shown with error bars. I show the reduced cumulants to remove the influence of the fitted convergence non-linear variance. This figure was published in **Barthelemy** et al. (2020a).

simulated PDFs and I already presented a way to circumvent these issues by means of the BNT transform (see section 1.9) that I will later apply to the convergence CGF and PDF.

As an illustration of the impact that the modelling errors on the CGF have on the cumulants I also show a comparison of the reduced third and fourth order cumulants (skewness and kurtosis) with their theoretical predictions in figure 3.2 for various redshifts and opening angles. A very good agreement is found in the weakly non-linear regime when the source redshift or the opening angle is large. As one goes towards a more non-linear regime, a clear departure is observed, the prediction systematically underestimating the measured skewness and kurtosis.

3.5. The convergence PDF

Formally, the convergence PDF is simply defined from the CGF as an inverse Laplace transform written as (more details in appendix A.1.2)

$$\mathcal{P}_\kappa(\kappa) = \int_{-i\infty}^{+i\infty} \frac{\mathrm{d}y}{2\pi i} \exp\left(-y\kappa + \phi_\kappa(y)\right). \tag{3.12}$$

However, the practical issue surrounding this integral comes from the fact that the CGF is only available in a numerical form for real values of y since the stationary condition – 3rd equation in (3.5) – is only solvable numerically for real y. More precisely, the matter rate function in equation (2.32) explicitly depends on the linear variance at a ρ – and thus y – dependent scale and thus the linear power spectrum which is only accessible numerically since no closed form is known for a ΛCDM universe. From there, I only know 3 directions that one could take to solve this issue. The first one involves performing a saddle point approximation which is not perfectly accurate but holds the advantage of being entirely

analytical. I personally never use this strategy for any comparison with simulation but it is still a quick way to access the PDF provided that the critical κ value is not too small. The second strategy, the one I use in this chapter, relies on finding a motivated analytical approximation of the linear variance that can then be continued in the complex plane, and the last strategy, that I use in the next chapter 4 for the aperture mass, relies on fitting an effective collapse that acts as a mapping between an effective un-smoothed Gaussian random field and the field whose PDF we aim at computing.

3.5.1 Saddle point (and beyond) approximation

The computation that appears in this subsection will also appear in a future paper Friedrich et al. (2021) in which I am a co-author.

Formally, the integration of (3.12) can be performed on any integration contour whose imaginary part begins at $-i\infty$ and ends at $+i\infty$. However numerically, it can be more efficient to choose it as the path that goes through the saddle point y^* with $\phi'(y^*) = \kappa$ and keeps the exponent $-y\kappa + \phi(y)$ real everywhere along the path (path of steepest decent, Valageas (2002)).

I here want to present an analytic approximation of this computation making use of the so-called Laplace's method as well as a next-to-leading order correction to it. Let me briefly re-explain its principle and consider integrals of the form

$$I_x = \int \mathrm{d}t \exp\left\{-\frac{R(t)}{x}\right\}, \tag{3.13}$$

where x is a small parameters and $R(t)$ is a function which has a global minimum at $t = t^*$. Without loss of generality, let me assume that $t^* = 0$. Then we have $R'(0) = 0$ and hence

$$\begin{aligned} R(t) =& R(0) + \frac{R''(0)}{2}t^2 + \mathcal{O}[t^3] \\ \equiv& R(0) + \frac{R''(0)}{2}t^2 + Q(t) \end{aligned} \tag{3.14}$$

such that the integral becomes

$$I_x = e^{-\frac{R(0)}{x}} \int \mathrm{d}t \exp\left\{-\frac{1}{2}\frac{t^2}{x/R''(0)}\right\} e^{-\frac{Q(t)}{x}}. \tag{3.15}$$

I then Taylor expand the exponential $e^{-\frac{Q(t)}{x}}$ up to t^6 which leads to

$$\begin{aligned} I_x \approx& e^{-\frac{R(0)}{x}} \int \mathrm{d}t e^{-\frac{1}{2}\frac{t^2}{x/R''(0)}} \left(1 - \frac{R'''(0)}{6x}t^3 - \frac{R^{(4)}(0)}{24x}t^4 + \frac{R'''(0)^2}{72x^2}t^6\right) \\ =& e^{-\frac{R(0)}{x}} \int \mathrm{d}t e^{-\frac{1}{2}\frac{t^2}{x/R''(0)}} \left(1 - \frac{R^{(4)}(0)}{24x}t^4 + \frac{R'''(0)^2}{72x^2}t^6\right) \\ =& \sqrt{\frac{2\pi x}{R''(0)}} e^{-\frac{R(0)}{x}} \left(1 - \frac{x}{8}\frac{R^{(4)}(0)}{R''(0)^2} + \frac{5x}{24}\frac{R'''(0)^2}{R''(0)^3}\right). \end{aligned} \tag{3.16}$$

Note that both the t^4 and t^6 terms of the expansion contribute at the same order in x, which is why we kept both of these terms. Laplace's method now consists in neglecting those contributions entirely and approximating

$$I_x \overset{\text{Laplace}}{=} \sqrt{\frac{2\pi x}{R''(0)}} e^{-\frac{R(0)}{x}}. \tag{3.17}$$

However it is possible to keep the next to leading order correction as done in equation (3.16).

Let me now re-formulate the PDF expression so that we can apply this approximation. We start by introducing a new variable $\lambda = \sigma^2(y - y^*)$ parametrising the deviation from the saddle point y^*, where $\sigma^2 = \langle\kappa^2\rangle_c$, such that

$$\mathcal{P}_\kappa(\kappa) = \int_{\tilde{\mathcal{C}}} \frac{\mathrm{d}\lambda}{2\pi\sigma^2 i} \exp\left\{\frac{-[\lambda^* + \lambda]\kappa + \sigma^2\phi([\lambda^* + \lambda]/\sigma^2)}{\sigma^2}\right\} \tag{3.18}$$

$$\equiv \int_{\tilde{\mathcal{C}}} \frac{\mathrm{d}\lambda}{2\pi\sigma^2 i} \exp\left\{\frac{-[\lambda^* + \lambda]\kappa + \varphi(\lambda^* + \lambda)}{\sigma^2}\right\}, \tag{3.19}$$

with $\lambda^* = \sigma^2 y^*$ and $\varphi(y) = \sigma^2\phi(y/\sigma^2)$ is the SCGF. Let us next introduce the function

$$S_\kappa(\lambda) = (\lambda^* + \lambda)\kappa - \varphi(\lambda^* + \lambda), \tag{3.20}$$

such that we now want to solve the integral

$$\mathcal{P}_\kappa(\kappa) = \int_{\tilde{\mathcal{C}}} \frac{\mathrm{d}\lambda}{2\pi\sigma^2 i} \exp\left\{-\frac{S_\kappa(\lambda)}{\sigma^2}\right\} \tag{3.21}$$

in the limit $\sigma^2 \to 0$. To do so, let me parametrise $\tilde{\mathcal{C}}$ as $\lambda(t) = it + f(t)$ with $\lambda(0) = 0$. Our integral then becomes

$$\begin{aligned}\mathcal{P}_\kappa(\kappa) &= \int_{-\infty}^{\infty} \frac{\mathrm{d}t}{2\pi\sigma^2}(1 - if'(t)) \exp\left\{-\frac{S_\kappa(it + f(t))}{\sigma^2}\right\} \\ &= \int_{-\infty}^{\infty} \frac{\mathrm{d}t}{2\pi\sigma^2} \exp\left\{-\frac{S_\delta(it + f(t))}{\sigma^2}\right\},\end{aligned}$$

where in the second step I used the fact that the exponent is real everywhere on the integration contour and that $\mathcal{P}_\kappa(\kappa)$ must be real. Using the fact that the path of steepest descent for the matter density PDF is symmetric around the real axis (Valageas, 2002) one can infer that $f'(0) = 0$ and $f'''(0) = 0$. Writing $f(t) \approx at^2$ and then demanding that the imaginary part of $S_\kappa(it + f(t))$ vanishes at $\mathcal{O}(t^3)$ one can furthermore show that

$$f''(0) = \frac{1}{3}\frac{S'''_\kappa(0)}{S''_\kappa(0)}. \tag{3.22}$$

Then defining $R(t) = S_\kappa(it + f(t))$ and using the 2nd order Laplace's method from above, I arrive at

$$\mathcal{P}_\kappa(\kappa) \approx \frac{e^{-R(0)/\sigma^2}}{\sqrt{2\pi\sigma^2 R''(0)}} \left(1 - \frac{\sigma^2}{8}\frac{R^{(4)}(0)}{R''(0)^2} + \frac{5\sigma^2}{24}\frac{R'''(0)^2}{R''(0)^3}\right) . \tag{3.23}$$

This expression finally simplifies because $R''(0) = -S''_\kappa(0) = \varphi''(\lambda^*)$ and the derivatives $R^{(n>2)}(0)$ involve the reduced cumulants encoded in $S_\kappa^{(n>2)}(0) = -\varphi^{(n)}(y^*)$ along with $f^{(k)}(0)$ where $k = 2\lfloor (n-1)/2 \rfloor$. Specifically, we have

$$R'''(0) = i\left(3f''(0)S''_\kappa(0) - S_\kappa^{(3)}(0)\right) = 0, \tag{3.24}$$

$$\begin{aligned} R^{(4)}(0) &= S_\kappa^{(4)}(0) - 6f''(0)S_\kappa^{(3)}(0) + 3f''(0)^2 S''_\kappa(0) \\ &= S_\kappa^{(4)}(0) - \frac{5}{3}\frac{S_\kappa^{(3)}(0)}{S''_\kappa(0)}, \end{aligned} \tag{3.25}$$

where the simplifications follow from the definition of $f''(0)$. Generally, this means that the higher-order corrections to the saddle-point are only relevant for a non-Gaussian PDF that reflect in a non-quadratic SCGF φ. In our case it seems that generally odd derivatives $R^{(2n+1)}(0) = 0$ vanish, as they would be purely imaginary and f is chosen to avoid that and furthermore, 2nd order saddle point approximation only requires $f''(0)$, while the next higher order would require $f^{(4)}(0)$.

Finally, the obtained expression is interesting in itself because it presents another formal expansion of the PDF in terms of its cumulants as is the case with the Gram-Charlier/Edgeworth expansion. It might be an interesting exercise to try and find if there are some formal links between the two, their relative convergence to the "true" PDF, etc.

3.5.2 Analytic continuation of the CGF

As mentioned above, the principal issue when performing the numerical integration (3.12) is the analytic continuation in the complex plane of the CGF which boils down to the analytic continuation of the linear variance of the matter density field. The linear variance in a long cylinder of Length L, radius $\chi\theta$ and at redshift z is defined as

$$\sigma^2_{\delta,l}(\chi\theta, L, z) = \int_0^\infty \frac{\mathrm{d}k_{||}}{2\pi} \int \frac{\mathrm{d}^2\boldsymbol{k}_\perp}{(2\pi)^2} P^l_\delta(k, z) W_{||}\left(k_{||}L\right)^2 W(\chi\theta k_\_)^2, \tag{3.26}$$

where $\boldsymbol{k}_\perp$ and $k_{||}$ are the components of the wave vector $\boldsymbol{k}$ orthogonal and parallel to the line-of-sight, and where $W_{||}$ and W are respectively longitudinal and transverse top-hat-windows. The radial component is of the order of $1/L$ and the transverse $1/D\theta$. Thus when L is large the radial part is negligible which leads to

$$\sigma^2_{\delta,l}(\chi\theta, L, z) = \frac{1}{L} \int \frac{\mathrm{d}^2\boldsymbol{k}_\perp}{(2\pi)^2} P^l_\delta(k_\perp, z) W(\chi\theta k_\perp)^2, \tag{3.27}$$

where $W(l) = 2J_1(l)/l$ and J_1 is the Bessel function of first order. Finally, notice that the $1/L$ term would as announced cancel out in the integration of the projected variance (3.3) and we thus get rid of it to only consider the variance of a 2D redshift slice of the matter density field. This expression is valid for any linear power spectrum and could in principle be computed analytically for a power spectrum with a closed form. For example with a power-law power spectrum,

$$P^l_\delta(k, z) = P_0(z)\left(\frac{k}{k_0}\right)^n, \tag{3.28}$$

the cylindrical variance reads

$$\sigma^2_{\delta,l}(\chi\theta, L, z) = \sigma^2_{\delta,l}(\chi_p\theta, L, z)\left(\frac{\chi}{\chi_p}\right)^{-n-2}, \tag{3.29}$$

where $\chi_p\theta$ is some pivot scale.

The power law approximation is unfortunately not accurate enough in our case since the lensing kernel integrates over many different scales but the last expression gives us a hint for a more sophisticated parametrisation for the variance that accounts for the running of the spectral index following Bernardeau et al. (2014a)

$$\sigma^2_{\delta,l}(\chi\theta, L, z) = \frac{2\,\sigma^2_{\delta,l}(\chi_p\theta, L, z)}{(\chi/\chi_p)^{n_1+2} + (\chi/\chi_p)^{n_2+2}}, \tag{3.30}$$

where $\chi_p\theta$ is still a pivot scale and n_1 and n_2 are parameters chosen to reproduce the correct variance by fitting the numerically integrated equation 3.27. Note that depending on the cosmology, n_1 and n_2 might also have a redshift dependence if the redshift dependence of the linear power spectrum cannot be factorised out with the linear growth factor as is the case for neutrino cosmologies. This does not concern us in this section but has its importance when evaluating the neutrino information contained in the PDF. A good fit that I use is for example $\chi_p = 4.5$ Mpc h^{-1}, $n_1 = -0.89$ and $n_2 = -1.97$. Note that the dependence on the cosmological parameters is entirely contained in the values of n_1 and n_2 and the pivot variance. This form thus ensures the analyticity of the variance mandatory for the continuation in the complex plane and one can obtain a linear variance precise up to a few percent across all needed scale within the line of sight integral.

Plugging this parametrisation of the linear variance into the matter rate function then allows one to solve the stationary condition for complex values of y, thus obtaining a complex matter and convergence CGF, and to perform the numerical inverse Laplace transform numerically, be it following the path of steepest method like in the previous subsection or with more brute force just using Simpson's method for numerical integration along the imaginary axis. The imaginary part of $\exp(-y\kappa+\phi_\kappa(y))$ oscillates quite a lot which necessitates a good sampling of the y values but the damping is fortunately fast enough that the integration still remains tractable.

3.5.3 The effective collapse

Another solution first used in Bernardeau & Valageas (2000) to obtain an analytic CGF consists in fitting an effective mapping, $\zeta(\tau_{\rm eff})$, between an effective un-smoothed Gaussian initial field and the lensing field (convergence or aperture mass) whose PDF we want to compute.

I start by re-writing the SCGF of the lensing field as given by the Legendre transform of the effective rate function

$$\varphi(\lambda) = \lambda\zeta(\tau_{\rm eff}) - \frac{1}{2}\tau^2_{\rm eff} \tag{3.31}$$

with the stationary condition written as

$$\lambda = \frac{\mathrm{d}}{\mathrm{d}\zeta}\frac{\tau^2_{\rm eff}}{2} = \tau_{\rm eff}\left(\frac{\mathrm{d}\zeta(\tau_{\rm eff})}{\mathrm{d}\tau_{\rm eff}}\right)^{-1}. \tag{3.32}$$

Note that now the effective mapping

$$\zeta(\tau_{\text{eff}}) = \sum_{k=0}^{n} \frac{\mu_k}{k!} \tau_{\text{eff}}^k, \tag{3.33}$$

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

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

$$\frac{\mathrm{d}\varphi(\lambda)}{\mathrm{d}\lambda} = \zeta(\tau_{\text{eff}}), \tag{3.34}$$

and thus from equation (3.31)

$$\frac{1}{2}\tau_{\text{eff}}^2 = \lambda \frac{\mathrm{d}\varphi(\lambda)}{\mathrm{d}\lambda} - \varphi(\lambda). \tag{3.35}$$

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

$$\mu_k^{\text{CGF}} = \sigma^k \, \mu_k, \tag{3.36}$$

and the inverse Laplace transform of the CGF can now be performed numerically – for example using Simpson's method along the imaginary axis – without any additional complication. This construction holds the advantage of also recreating the presence of critical points (non-convexity of the rate function or equivalently multiple solutions to the stationary condition) along the real axis. However, due to the (high-order) polynomial nature of the effective mapping, there are now two critical points, one negative and one positive which makes this construction, at least on paper, more appropriate for the aperture mass than for the convergence for which there is only one positive critical point along the real axis. This approach might still be useful for the convergence PDF but only making sure that the negative critical point is far enough from the origin so that the negative κ tail is not too much affected.

Additionally, from a mathematical point of view, this construction is strictly valid for quantities satisfying a large deviation principle such as the density field in thin redshift slices along the line of sight. This in particular means that this effective procedure fails, at least on paper, to convey some of the information that the large-deviation approach to the projected CGF does encode. Though this is not reflected in the successive cumulants of the field which are well reproduced by the effective approach, this is the case in the tails of the PDF or equivalently in the CGF near the critical points. Thus a more tedious but more appropriate effective approach to the analytical lensing CGF would rather be to fit an effective mapping at the level of each redshift slice along the line of sight. Fortunately, I do

not in practice find any significant difference in the PDF tails between fitting an effective collapse at each slice or directly on the projected CGF, which is explained by the fact that the asymptotic result is reached rather far in the tails (Bernardeau et al., 2014a) and thus the main contributions to the shape of the PDF come from the few first cumulants.

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

Finally, let me emphasise that effectively, only the knowledge of the few first cumulants allows us to recover the entire PDF. This is as we saw enforced by the formal construction obtained through assumption that the convergence and aperture mass satisfy – this is only true with the effective mapping – a large deviation principle, and also by the fact that we impose our rate function to be convex through the use of the Legendre rather than Legendre-Fenchel transform, which effectively acts as if the rate function was linear beyond the critical points. The unchanging convexity of the rate function imposes that the PDF is uni-modal. This can be for example put into perspective with works in the field of mathematics, see for example Gavriliadis & Athanassoulis (2009), that present, in general, how much of the PDF can be recovered just from the information of a certain number of moments. It turns out that uni-modal PDFs can be very well approximated by a formal and general reconstruction coming from their first few cumulants. This partly explains why the effective mapping approach can work to such precision given that more constrains apply.

I will give a more quantitative assessment of how well this approaches reproduces the CGF when using it for the computation of the aperture mass PDF in the next chapter 4.

I now conclude this section by giving the link that can be made between the μ coefficients and the cumulants of the field. Given that the fitted μ values and the computed cumulants agree very well, computing cumulants to obtain those coefficients instead of fitting them might become a viable option, especially in cases where the large deviation formalism is used for filters other than top-hats and where imposing the stationary condition even to obtain the real space CGF is a numerical challenge in itself (Reimberg & Bernardeau, 2018). Since the reduced cumulants of the field are given by the successive derivatives of $\varphi(\lambda)$ taken in zero, I obtain

$$S_{3,M_{\rm ap}} = 3\mu_2, \tag{3.37}$$
$$S_{4,M_{\rm ap}} = 12\mu_2^2 + 4\mu_3, \tag{3.38}$$
$$S_{5,M_{\rm ap}} = 60\mu_2^3 + 60\mu_2\mu_3 + 5\mu_4, \tag{3.39}$$
$$S_{6,M_{\rm ap}} = 360\mu_2^4 + 720\mu_2^2\mu_3 + 90\mu_3^2 + 120\mu_2\mu_4 + 6\mu_5. \tag{3.40}$$

θ (arcmin)	10	20	30	40	50
σ^2_κ (10^{-5})	9.1	5.4	3.7	2.8	2.3

Table 3.1: Variance of the convergence field fitted from the simulated data for various opening angles as labelled and associated with figure 3.3. The source redshift is $z_s = 1.5$.

These relations can be inverted and one finally arrives at

$$\mu_2 = \frac{S_3}{3}, \tag{3.41}$$

$$\mu_3 = \frac{-4S_3^2 + 3S_4}{12}, \tag{3.42}$$

$$\mu_4 = \frac{40S_3^3 - 45S_3S_4 + 9S_5}{45}, \tag{3.43}$$

$$\mu_5 = \frac{-560S_3^4 + 840S_3^2S_4 - 135S_4^2 - 192S_3S_5 + 24S_6}{144}. \tag{3.44}$$

3.6. Theoretical versus simulated PDFs

For a comparison of the now-established theoretical model for the convergence PDF with the Takahashi simulation, I choose the analytic continuation of the CGF method that I just described in section 3.5.2. I also choose to model the driving parameter along the line of sight with the pivot-scale-like approximation also presented above in equation (3.30) following the discussion of section 3.3.2. As a consequence, the only free (but physical) parameter of the theory is the non-linear variance of the convergence field that can be directly measured from data or the simulation, predicted from another theoretical formalism (Halofit, the Euclid emulator or even 2-loop power spectrum), or even fitted in the data/simulation so that the PDF is the most accurate possible in the bulk where the influence of the variance is most notable.

I now show in figure 3.3, for a fixed source plane at $z_s = 1.5$ and for different opening angles from $\theta = 10$ to 50 arcmin, our theoretical PDFs compared to the ones measured in the simulation. The upper panels show the PDFs with a linear scale to emphasize the behaviour around the maximum of the PDF and with a log-scale to highlight the exponential decay in the tails of the distributions suggested by the formalism. In addition, the lower panel shows the relative difference between the theoretical and measured PDF in the 3 σ-region around the maximum. As for the comparison between the theoretical and measured CGFs, I only use one full-sky realisation of the Takahashi simulation and the error bars are the error on the mean among 8 sub-samples of the full-sky. I also give in table 3.1 the result of the different measured variances.

Alternatively, figure 3.4 shows the same comparison when fixing the opening angle and varying the redshift of the source plane with the fitted variances in table 3.2.

In both cases we observe that as expected, the theoretical PDF becomes more and more accurate as one approaches the linear regime for a rising part of the scales and redshifts of the probed underlying density field (the $\sigma_\delta \to 0$ limit for higher and higher redshifts at a

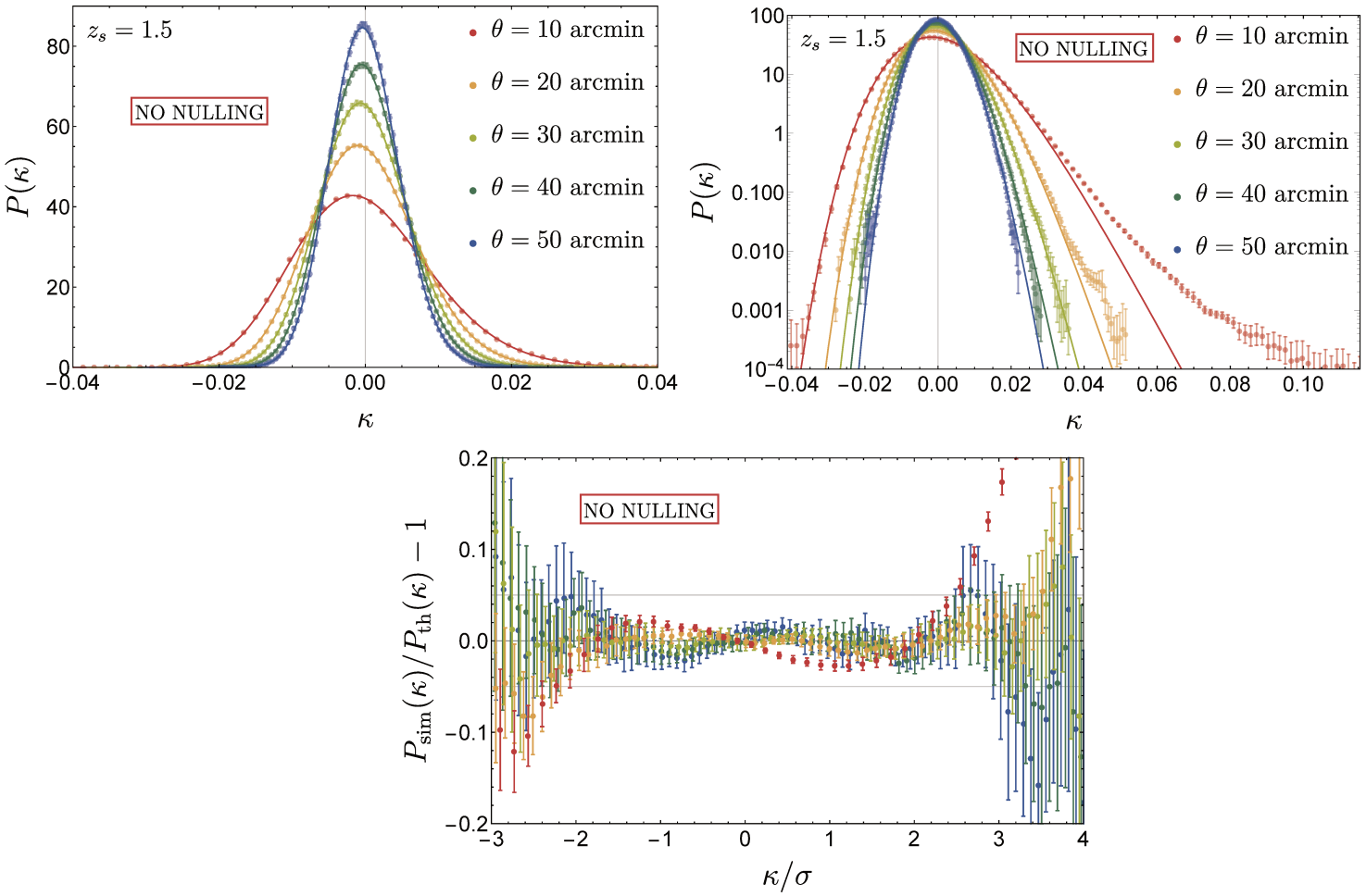


Figure 3.3: One point PDF of the weak-lensing convergence for different opening angles from 10 arcmin (red) to 50 arcmin (blue) as labelled. The source redshift is fixed here to $z_s = 1.5$. Solid lines display the theoretical predictions given by the integration along the line of sight of the large deviation formalism while the measurements on the simulated sky (Takahashi's simulation) are shown with error bars. Top left panel: PDF in linear scale. Top right panel: Same as top left panel in log scale to better display the tails. Bottom panel: residuals of the simulated data compared to the prediction. This figure was published in **Barthelemy et al. (2020a)**.

z_s	1	2	3.1	4.2	5.3
σ^2_κ (10^{-4})	0.47	1.4	2.3	3.1	3.7

Table 3.2: Variance of the convergence field fitted from the simulated data for various source redshifts as labelled and associated with figure 3.4. The opening angle is $\theta = 10$ arcmin.

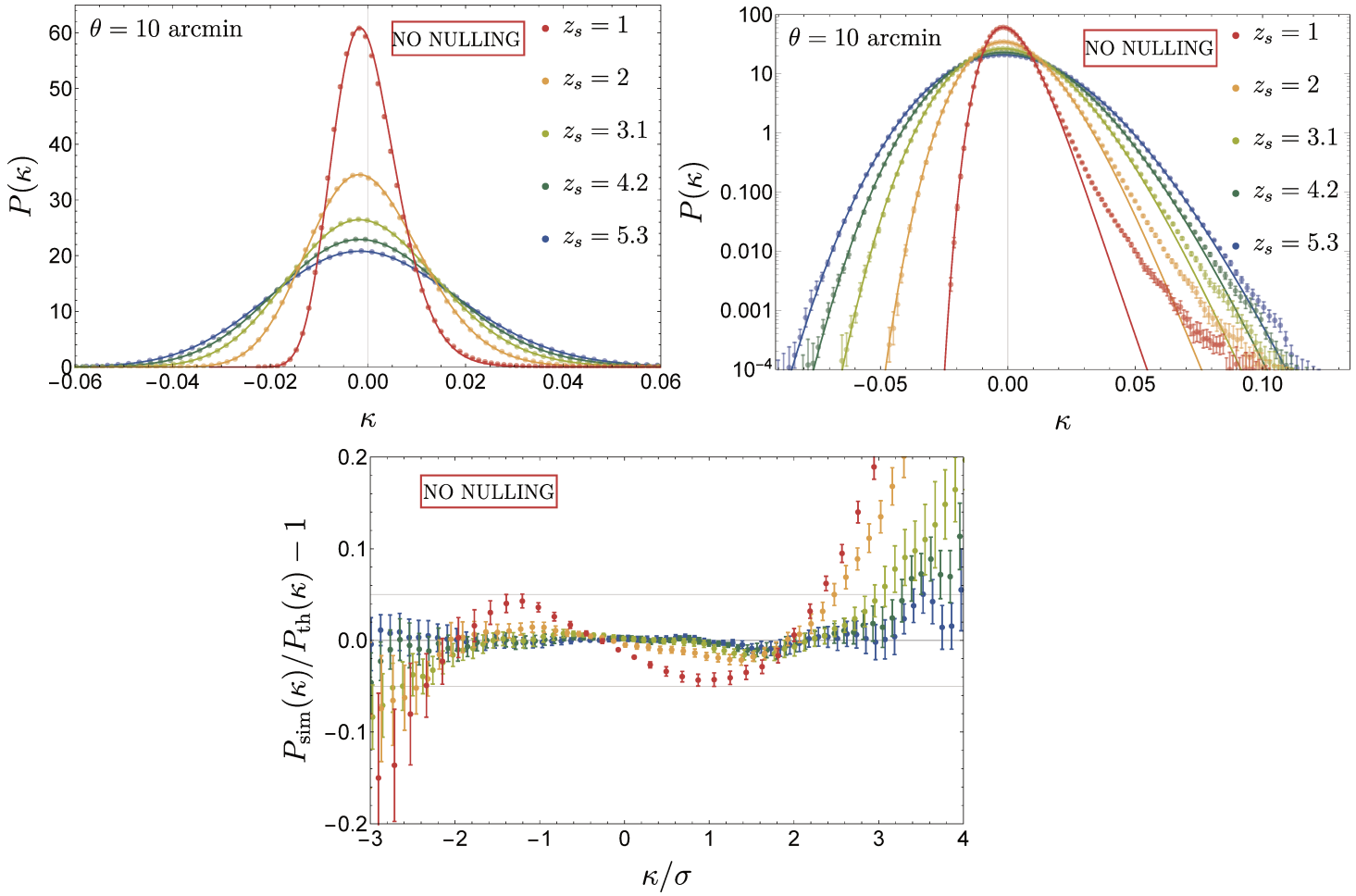


Figure 3.4: One point PDF of the weak-lensing convergence for different redshifts from 1 (red) to 5.3 (blue) as labelled. The opening angle is here fixed to $\theta = 10$ arcmin. Solid lines display the theoretical predictions while the measurements on the simulated (Takahashi's simulation) sky are shown with error bars. Top left panel: PDF in linear scale. Top right panel: Same as top panel in log scale to better display the tails. Bottom panel: residuals of the simulated data compared to the prediction. This figure was published in **Barthelemy** et al. (2020a).

fixed angular scale or larger angular scales at fixed source redshift). Still, in this regime (towards the blue curves) the LDT prediction provides us with a better description than the linear Gaussian case, especially in the tails where the departure from a pure Gaussian is clearly seen and well reproduced by our theory. When diving into a more non-linear regime (towards the red curves), the distribution clearly gets more skewed towards low convergences, and the LDT prediction captures relatively well this non-linear evolution and remains within 5 per cent from the measured distribution in the 2-σ region around the mean convergence. However, diving further away in the (rare event) tails, the agreement between our prediction and the simulation gets worse as expected since the spherical collapse dynamics becomes less and less accurate while the influence from the very small scales at the tip of the light-cone becomes very noticeable. This is clearly seen on the residuals where a typical H_{e3} modulation by the skewness is visible, showing that higher order correction to the skewness becomes necessary. Indeed, let me reiterate here that the skewness enters the Edgeworth expansion (see also appendix A.1.3) of the PDF at the first non-Gaussian correction order and multiplies a third order Hermite polynomial of the convergence field as follows

$$\mathcal{P}(\kappa) = \mathcal{G}(\kappa)\left[1 + \sigma\frac{S_{3,\kappa}}{3!}H_{e3}\left(\frac{\kappa}{\sigma}\right) + \mathcal{O}(\sigma^2)\right], \tag{3.45}$$

where $H_{e3}(x) = x^3 - 3x$. Though useful to interpret the shape of the residuals, let me however emphasise that the prediction presented here is based on LDT and does not use or assume a Edgeworth expansion. Interestingly though, one can show that our approach is equivalent to having an infinite Edgeworth series (Bernardeau & Kofman, 1995) – that is to say with no truncation – but with reduced cumulants given by spherical collapse along the line of sight. This infinite series is one of the advantages of the LDT based formalism as it allows us to get accurate (and physical) predictions for the tails of the PDF. This is to be contrasted with a truncation at a given order in the Edgeworth expansion which by construction – if it captures correctly the vicinity of the maximum – would get very inaccurate and nonphysical in the tails (the truncated PDF becoming negative for some values of the convergence and not normalised).

Overall, it is found that – unsurprisingly – theoretical predictions of the CGF and PDF are valid in a somewhat narrower regime than the same formalism for the 3D densities: while the region around the most probable value is quite well captured, the high convergence tails are heavily affected. As previously noted, a large part of this limitation comes from the lensing kernel that mixes all scales and in particular the small ones at the tip of the light-cone that cannot be well modelled by the spherical collapse dynamics. One way to circumvent this issue is obviously the so-called BNT transform as I will later demonstrate in this chapter in section 3.9. For now I will move to standard extensions of this model and focus on the cosmological information content of the correctly modelled bulk of the PDF.

3.7. Additional corrections to the convergence PDF

3.7.1 Shape noise

The addition of shape noise onto our theoretical model for the convergence PDF is particularly straightforward if we, following the literature, still model it as an independent random

Gaussian noise on top of the lensing signal. The noisy convergence PDF is then given by the convolution of our PDF with a Gaussian of appropriate variance given in equation (1.46)

$$\mathcal{P}_{SN}(\kappa) = \frac{1}{\sqrt{2\pi}\sigma_{SN}} \int_{\kappa_{\min}}^{\infty} d\kappa' \exp\left(-\frac{(\kappa-\kappa')^2}{2\sigma_{SN}^2}\right) \mathcal{P}(\kappa'). \tag{3.46}$$

In a Euclid-like survey where one expects to observe a galaxy density of 3 arcmin^{-1} in each equally populated bin – thus source plane in this simplified case – and for a smoothing angle of 10 arcmin, I thus expect $\sigma_{SN}^2 \sim 3 \cdot 10^{-4}$ and less with increasing smoothing angle thus roughly doubling the variance of the measured signal if I refer to tables 3.1 and 3.2. The question arising in this context is thus whether or not non-Gaussianities remain detectable, that is whether or not the PDF remains an interesting probe. To get a rough idea of the answer let me focus a bit on the skewness and assume that the size of its experimental error bar is given by the rms of its estimator.

In practice the 3$^{\text{rd}}$ cumulant of the field is measured by an estimator that we choose here to be given by the so-called k-statistics. Thus defining the sums of the r^{th} powers of the n independent data points (seen as n effective independent realisations of the κ values denoted X_i) as

$$s_r \equiv \sum_{i=1}^{n} X_i^r, \tag{3.47}$$

an unbiased estimator of $\langle\kappa^3\rangle_c$ that we call k_3 is given by

$$k_3 = \frac{n^2 s_3 - 3n s_2 s_1 + 2s_1^3}{(n-2)(n-1)n}. \tag{3.48}$$

We can thus compute* the variance of the estimator, $\sigma_{k_3}^2$,

$$\left\langle \left(k_3 - \langle\kappa^3\rangle_c\right)^2 \right\rangle = \frac{6\langle\kappa^2\rangle_c^3 n}{(n-2)(n-1)} + \frac{9\langle\kappa^4\rangle_c\langle\kappa^2\rangle_c}{n-1} + \frac{9\langle\kappa^3\rangle_c^2}{n-1} + \frac{\langle\kappa^6\rangle_c}{n}, \tag{3.49}$$

which can then be used to estimate the error bar on the measured cumulant.

As already seen, in the presence of Gaussian shape noise considered statistically independent from the κ field, only $\langle\kappa^2\rangle_c$ is affected and one would only need to replace $\langle\kappa^2\rangle_c^3 n$ in equation (3.49) by

$$\langle\kappa^2\rangle_c \to \langle\kappa^2\rangle_c + \sigma_{SN}^2.$$

The form given in equation (3.49) thus renders apparent how the number of data points, the shape noise and the amplitude of the signal which depends on the scales and redshifts probed will make the measurement of $\langle\kappa^3\rangle_c$ compatible with zero, that is whether or not non-Gaussian features will be detected in the PDF. There only remains to estimate the number of data points that are independent. Though not exact, since the correlation between disks will rapidly decay as their distance grows, we can put an upper limit on the number of data points by computing the number of non-overlapping disks one can draw on the surface

*This is easily done by hand in a few lines but here is an appropriate place to highly recommend the mathStatica package for Mathematica (Wolfram, 2021) which does many exact statistics computation.

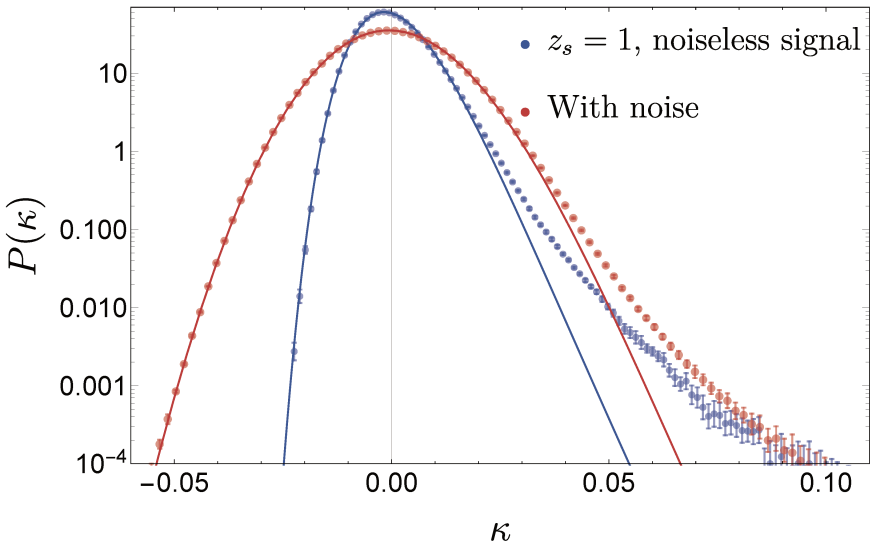


Figure 3.5: Effect of shape noise on the convergence PDF for an opening angle of 10 arcmin at source redshift $z = 1$. Points with error bars are measured in the Takahashi simulation while the solid lines are theoretical predictions convolved with a Gaussian where relevant. This figure was published in **Barthelemy** et al. (2020a).

area. The error bar computed in this fashion can then be shown to be very close to the ones estimated from the numerical simulation from the standard deviation on the skewness among different and independent full-sky realisations. In our example, using a very rough numerical solution to the Thompson problem and the Euclid experiment approximately covering 36% of the full-sky, I find that 132 500 independent disks of 10 arcmin can be drawn. Using the theoretical values for the convergence at redshift $z_s = 1$ and the amplitude of shape noise I find that $k_3 = 2.7 \cdot 10^{-7}$ and $\sigma_{k_3} = 4.6 \cdot 10^{-8}$ which renders detectable the presence of non-Gaussianities even in the presence of shape noise! Still for the purpose of illustration, let me also show in figure 3.5 the convergence PDF with and without shape noise at this redshift and angular scale where I also added noise in the simulated map by addition of a random Gaussian term of the appropriate amplitude in each simulated pixel.

3.7.2 Post-Born corrections

Chapter 1 introduced several additional corrections, especially to the convergence and aperture mass skewness. Through the prism of post-Born corrections, I now want to incorporate those corrections into the one-point PDF of the convergence field. The strategy to adopt in this respect is not trivial and I want here to present three different approaches that may each work in different contexts, though only the 2$^{\text{nd}}$ and 3$^{\text{rd}}$ one are mathematically rigorous and only the 3$^{\text{rd}}$ one is appropriate in the context of post-Born corrections.

The fast and dirty approach: Remember that the SCGF whose successive derivatives in zero define the scaled cumulants such as the skewness S_3 and the kurtosis S_4 can be

expressed as

$$\varphi_\kappa(y) = \sum_{n=0}^{+\infty} S_n \frac{y^n}{n!}. \tag{3.50}$$

Since this object is directly accessible in our formalism, and given that we only want to modify S_3 for post-Born effects*, one attractive option could be to add the post-Born correction by hand directly in the SCGF with the following prescription

$$\varphi_\kappa^{\rm corr}(y) = \sum_{n=0}^{+\infty} S_n \frac{y^n}{n!} + S_{3,\kappa}^{\rm corr} \frac{y^3}{6}. \tag{3.51}$$

This strategy is found to work well for small corrections to the skewness (as is the case when including primordial non-Gaussianities for instance, see Friedrich et al. (2019)). However this is a shaky construction from a mathematical point of view since there is absolutely no guarantee that the newly defined sequence of cumulants actually corresponds to a random variable, in other words that this new function has all the good properties of a SCGF, in particular to be a PDF once inverse Laplace transformed (positive, normalised, etc). This issue is known as the *cumulant problem*, which states that there is currently no known set of (in)equalities that constrain how any infinite † sequence of real numbers can represent the sequence of cumulants of a random variable ($11^{\rm th}$ problem in Rota (2001)).

This can also be hinted using standard results of perturbation theory by showing that the cumulants are all inter-dependent and thus one cannot simply modify one without touching consistently the others. Indeed it was shown (Bernardeau, 1995; Bernardeau et al., 2002) that the reduced cumulants S_n as given by the spherical collapse model are all expressed as a function of the so-called vertices ν_i of order $i = n$ *and less*. These vertices appear in the spherical collapse model as the Taylor coefficients of the collapse dynamics seen as a mapping between the non-linear (late-time) density and the initial density contrast (see appendix A.3.2). For example the skewness and kurtosis at tree order read‡

$$\begin{aligned} S_3^{2D}(R) &= 3\nu_2 + \frac{3}{2}\frac{{\rm d}\log(\sigma^2(R))}{{\rm d}\log(R)}, \\ S_4^{2D}(R) &= 4\nu_3 + 12\nu_2^2 + (14\nu_2 - 2)\frac{3}{2}\frac{{\rm d}\log(\sigma^2(R))}{{\rm d}\log(R)} \\ &\quad + \frac{21}{4}\left(\frac{{\rm d}\log(\sigma^2(R))}{{\rm d}\log(R)}\right)^2 + \frac{{\rm d}^2\log(\sigma^2(R))}{{\rm d}\log(R)^2}, \end{aligned} \tag{3.52}$$

in a 2D slice of radius R and where the values of the vertices ν_n are tabulated for the exact 2D spherical collapse ($\nu_2 = 12/7$, $\nu_3 = 29/7$, etc). This illustrates the need for a joint

*It would be more correct to say that we only know the correction to S_3 or more generally only know a finite set of corrections to this set of cumulants, the problem being the exact same.

†A sequence of two numbers, the second being strictly positive, always corresponds to a Gaussian, otherwise a cumulant generating function must contain an infinite number of non zero cumulants, the variance always being a strictly positive value.

‡Note that this is strictly valid only in the case of an Einstein-de-Sitter background but found to be an excellent approximation for standard FLRW cosmology.

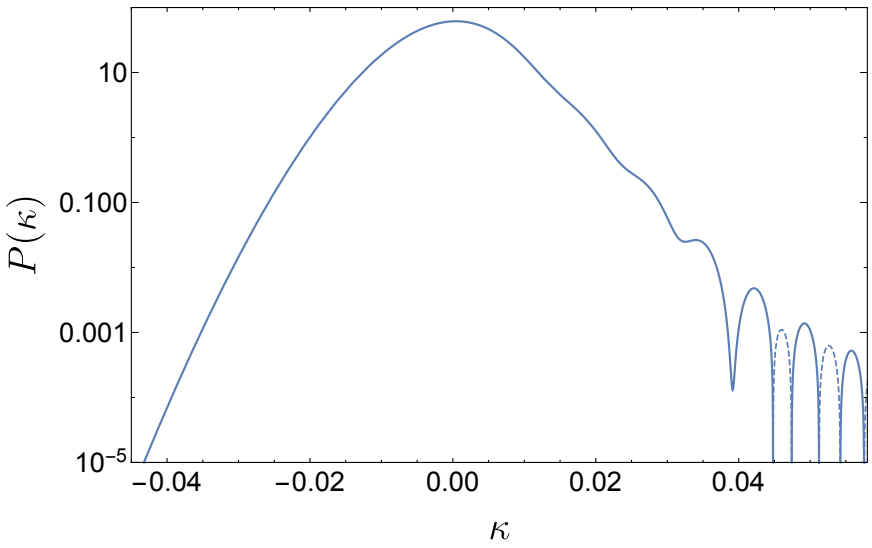


Figure 3.6: Example of an ad-hoc PDF obtained "by hand" after a modification of the skewness using the fast and dirty approach, that is by modifying the cumulant generating function directly. The solid line shows the positive values of the PDF while dashed lines represent the negative values. It is clear that the resulting function displayed on this figure does not possess the elementary mathematical properties of a PDF such as being positive. The highly oscillatory behaviour in the tails is also un-physical unlike what can be obtained using our theoretical formalism. This figure was published in **Barthelemy** et al. (2020b).

modification of *all* cumulants simultaneously in order to get a consistent modelling of the full one-point statistics. If one changes the value of the skewness (to correct for the post-Born effects for instance), one could change accordingly the value of the other cumulants so as to keep the interpretation in terms of an underlying mapping (or vertex structure). Let us finally show a concrete example of the issue with figure 3.6 which displays the result one would obtain when trying to alter the skewness in an arbitrary PDF* using this approach. It is clear that this function, although each successive integral yields the correct values of the cumulants (this is actually used to check that the inverse Laplace transform was performed successfully), cannot be considered a correct PDF since it shows negative values and a highly oscillating behaviour in the tails. As a consequence, since post-Born corrections are quite important at high source redshifts compared to the non-linearities generated by the density field, I do not use this approach.

Post-Born effect on each lens: What actually ensures that our matter PDFs are indeed PDFs is the formal construction from large deviation theory where all the cumulants are obtained from a "most probable" mapping between the initial conditions of the field and the field at late times. Our approach to the convergence PDF thus in some sense also consists in the description of an implicit mapping between some initial field and the final field where the mapping is built from the dynamics of the underlying density field. Therefore, since taking into account post-Born corrections must in a sense be equivalent to finding a more appropriate mapping for the convergence, we can change the mapping of the underlying

*Here, we choose the PDF of the convergence for a source redshift $z_s = 1$ and aperture $\theta = 10$ arcmin for which we subtracted 70 to the original skewness of 77.

density field to ensure that it effectively encompasses those corrections for the convergence.

One idea is to modify, for each density slice that makes the convergence, the value of the 2D spherical collapse parameter ν so as to reproduce the corrected convergence skewness (tree-order correction + spherical collapse) instead of the uncorrected one. In physical terms, changing the mapping along the line-of-sight would account for the fact that close lenses act as a deformation of background lenses and thus alter the collapse the observer sees. Indeed, the correction displayed in equation (1.83) can be seen as a line-of-sight correction to the third cumulant of each density slice and thus a correction of the collapse parameter ν. Thus if one sees the correction as the integral of some function $F(z, z_s)$,

$$\left\langle \kappa^3_{\rm corr} \right\rangle_\theta = \int_0^{z_s} \frac{\mathrm{d}z}{H(z)} F(z, z_s), \tag{3.53}$$

and uses the projection formula (3.4) to get the uncorrected tree-order skewness, one can write the corrected convergence as

$$\left\langle \kappa^3 \right\rangle_\theta = \int_0^{z_s} \frac{\mathrm{d}z}{H(z)} \omega(z, z_s)^3 \left(\left\langle \delta^3 \right\rangle + \frac{F(z, z_s)}{\omega(z, z_s)^3} \right), \tag{3.54}$$

where $(\langle \delta^3 \rangle + F(z, z_s)/\omega(z, z_s)^3)$ will lead to the new skewness value of the slice and hence the corrected collapse parameter. Note that since $F(z, z_s)$ is itself an integral from the observer to the slice it indeed states how foreground lenses affect each density slice. However for the mapping to make sense in each slice along the line-of-sight, that is keeping the functional form of the spherical collapse, it is mandatory that the collapse dynamics still dominates over the post-Born corrections, in other words that the contribution to the convergence skewness coming from the gravitational collapse in each slice is more important than the one coming from the integrand of equation (3.53). Indeed, it would not make sense to keep the functional form describing the collapse dynamics if it is not the dominant effect anymore.

I plot in figure 3.7 the contribution to the skewness of each slice depending on the redshift of the slice for a source redshift $z_s = 1100$ and an opening angle $\theta = 10$ arcmin. This plot shows that the contribution from the post-Born effects, although overall subdominant, dominates the dynamics of the slices for a large fraction of the line-of-sight. This leads to negative values of ν in many slices which totally changes the shape of the mapping. As such, modifying the spherical collapse mapping for all the slices would not make any sense and would lead, from a physical point of view, to an overall bad construction.

A large deviationesque construction: As was discussed in the previous paragraph, the physics of the large deviation approach of cosmic structure formation is entirely contained in the mapping from the field initial conditions to its late-time counterpart and thus any modification that one would want to incorporate should only be at the level of the mapping itself. I therefore stick to this strategy. Note that this modification of the mapping should be done with care so as not to introduce any spurious critical points where the new rate function ceases to be convex – in the sense that they would for example emerge from a truncated mapping series expansion – in which case the asymptotic tails of the PDF could be affected because of the presence of unwanted critical points.

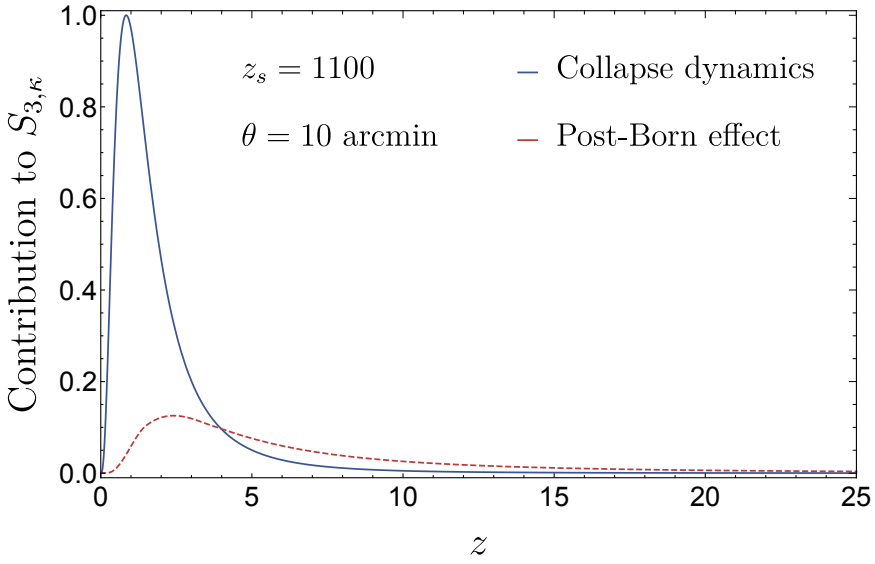


Figure 3.7: Absolute contributions to the convergence skewness coming from each slice and from the 2D spherical collapse dynamics (blue) and the post-Born corrections (red). Dashed line represent negative values. Here the contributions are shown for a source redshift of $z_s = 1100$ and an opening angle of $\theta = 10$ arcmin. This figure was published in **Barthelemy** et al. (2020b).

Since the effect of the post-Born corrections on the skewness is subdominant for the convergence with respect to the total collapse dynamics in the cone, we might thus want to consider a modification of the implicit "conical" collapse driven by the parameter ν. In this case, we keep the functional form of the 2D spherical collapse for each slice and find an overall effective ν modified for each of them. For a source redshift $z_s = 1100$ and an opening angle $\theta = 10$ arcmin for which the post-Born corrections are important, I plot in figure 3.8 how the effective dynamics are affected. This change of effective mapping will also impact the value of the higher order cumulants (kurtosis and beyond) but in a meaningful, or at least mathematically consistent, way since they are now deduced at the level of the new mapping and so that issues (negative values, mis-normalisation) described for the fast and dirty approached are now avoided. Moreover, since this new construction also respects the cumulant hierarchy imposed by the functional form of the mapping we can safely expect that the higher-order cumulants will at least have the good order of magnitude except maybe for highly non-linear regimes if this hierarchy does not apply.

Application to CMB lensing: I showed in figures 1.3 and 1.4 that the post-Born corrections in the context of weak lensing experiments with sources of low redshift, $z_s \leq 3-4$, will not be of the utmost importance and we have seen when comparing our formalism for the PDF without those corrections to the Takahashi simulation that their absence do not impact the quality of the prediction. However, since the value of the tree-order correction to the skewness is nearly constant and close to $S_{3,\kappa}^{\rm corr} \approx -1$ and the (uncorrected) convergence skewness roughly decreases as $\approx z_s^{-1.35}$, depending on the precision one may need there is always a regime for which it might be mandatory to include the correction to the PDF. A good example is to focus on CMB lensing ($z_s = 1100$) where the correction is of the order

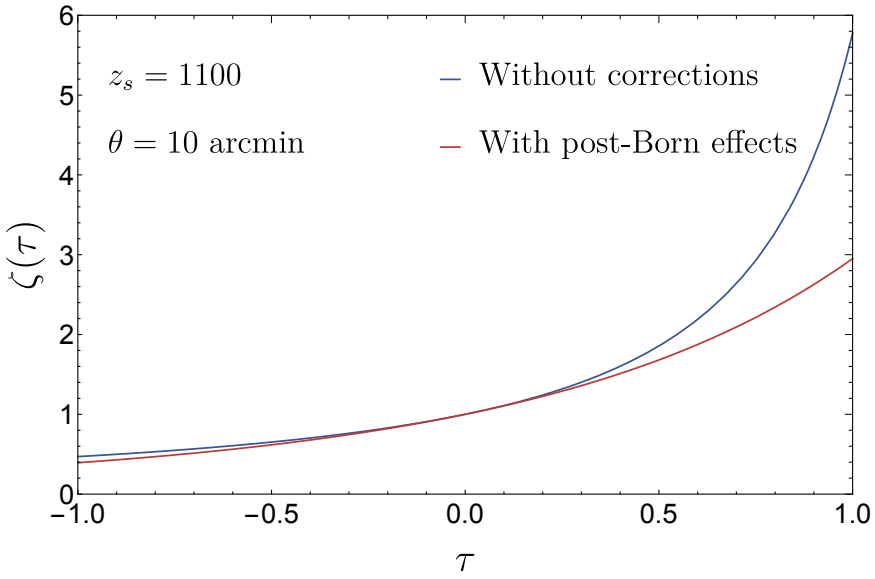


Figure 3.8: Modified mapping of a slice to account for post-Born corrections to the CMB convergence skewness (red) compared to the uncorrected spherical collapse dynamics (blue). The corrections are introduced by changing the mapping globally and for a source redshift $z_s = 1100$ and an opening angle $\theta = 10$ arcmin (ν goes from 1.4 to 6.7). This figure was published in **Barthelemy** et al. (2020b).

of the uncorrected skewness and thus plays an important role. Following the conclusion of the previous paragraph, let me globally modify the spherical collapse parameter so as to reproduce the corrected convergence skewness, plug this new parametrisation into our convergence PDF formalism and thus compute the one-point PDF of the convergence field taking into account the post-Born corrections. I then compare in figure 3.9 the computed PDF at source redshift 1100 for an opening angle of $\theta = 10$ arcmin to the one extracted from 23 realisations of the Takahashi simulated maps (for which the measured standard deviation is $\sigma_\kappa \sim 0.036$).

The first thing worth noting is that the simulated cosmic variance is relatively large as can be seen from the size of the error bars for both the tails of the PDF and the higher-order cumulants on their own*. Nevertheless the need for going beyond the Born-approximation is clear when one looks at the residuals of figure 3.9 (red points) that shows significant (tens of percent) deviations between the simulation and the theoretical prediction without post-Born corrections, manifestly driven by an order one error on the skewness (as the S-shape suggests, being a well-known H_{e3} modulation). This observation strengthens my previous argument stating that taking into account the correction to the skewness is sufficient to get an accurate PDF. Once post-Born corrections are accounted for (blue solid line), the agreement between the theoretical formalism and the measurement in the simulation is very good within the error bars in the range of convergence probed, that is to say in the $\sim 5\sigma_\kappa$ region around the mean. For negative values of the convergence, there is no sign of any deviation from the prediction and a subpercent accuracy is reached. On the other hand, some hints of departures are seen in the high (positive) convergence tail for which more

*Note that large-deviation theory provides us with a way to compute the expected error bars including both the effect of shot noise and cosmic variance, we refer the reader to Codis et al. (2016) for more details.

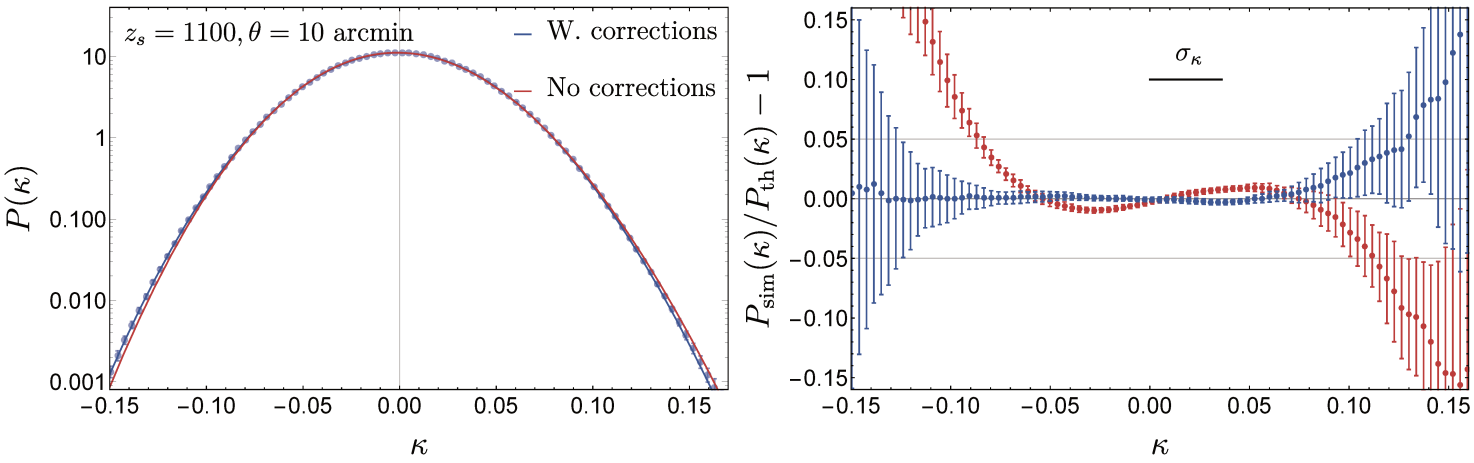


Figure 3.9: One point PDF of the weak lensing convergence for an opening angle of 10 arcmin as labelled. The source redshift is fixed here to $z_s = 1100$. Solid lines display the LDT predictions with (blue) and without (red) implementation of post-Born corrections while the measurements on the simulated sky is shown with error bars obtained from the standard deviation between 23 realisations. Top panel: PDF in log scale to better display the tails. Bottom panel: residuals of the simulated data compared to the prediction. This figure was published in **Barthelemy** et al. (2020b).

accurate simulations would be needed (at least for κ above $\sim 2\sigma_\kappa$).

Overall, it is very reassuring and satisfying to see that theoretical predictions and simulations agree within a percent in a large range of values for the convergence, from $\sim -5\sigma_\kappa$ to $2\sigma_\kappa$, a regime where both theory and simulations have not been tested in detail. For higher values of the convergence field, it is difficult to draw any conclusion at this stage with regards to either the accuracy of the prediction or the validity of the simulated maps. Indeed though it is clear that the theory tends to systematically under-predict the expectation values of the PDF in the high tail, an effect which comes from the fact that the skewness itself is under-predicted with respect to its simulated expectation value, given the quite sizeable dispersion (cosmic variance displayed as the error bars) one gets between the realisations, the fact that the simulation was not extensively tested with non-Gaussian statistics or the fact that full-sky CMB lensing simulation is still something rather hard to perform we cannot exclude that some systematic comes from the simulation, nonetheless being state-of-the-art, itself. In any case this construction boils down to constructing a "spherical collapse" convergence PDF with some post-Born correction, it is thus not excluded that one might in the future have to go to higher order to get a more accurate prediction if needed.

Let me conclude this part about post-Born corrections on the PDF by having a look at the kurtosis. Figure 3.10 displays the measurements from the simulations against the LDT prediction without post-Born corrections (red). In addition, we overlay in blue the value of the kurtosis in the formalism developed in this subsection, that is to say when post-Born corrections *to the skewness* are accounted for by modifying the collapse parameter in the large-deviation formalism. Hence, the red line in Fig. 3.10 comes from a physical formalism - large deviation theory with a meaningful 2D spherical collapse - but without any post-Born corrections while the blue-line is only a consequence of a correction of the spherical collapse parameter only meaningful for the skewness. Interestingly, the resulting correction to the

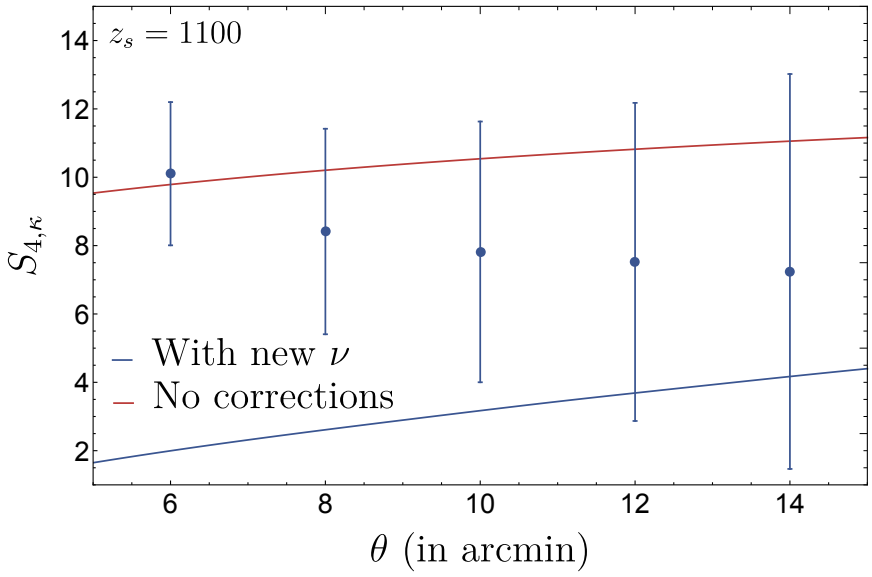


Figure 3.10: Convergence kurtosis of the CMB with respect to the opening angle of the top-hat smoothing applied. In red is the kurtosis evolution as computed from large deviation theory without taking into account post-Born corrections while in blue these corrections are taken into account via a modification of the collapse parameter to match the corrected skewness. The measured points are from 23 Takahashi's simulated maps (mean of the 23 realisations) and error bars represent their standard deviation around the mean. This figure was published in **Barthelemy** et al. (2020b).

kurtosis does not lead to completely unrealistic values (which could have been a real worry and would have made our model unworkable, notably in the tails) and even seems to go in the right direction, lowering the uncorrected value as it should. Given the quite large error bars on the kurtosis, it is difficult to quantify precisely the error we make on this cumulant (do we tend to slightly underestimate it for instance?). At the level of accuracy of current state-of-the-art CMB lensing simulations, such an effect is completely within the error bars but it would be nonetheless interesting to compare to the tree order kurtosis with post-Born corrections that one could obtain similarly to the skewness but going to next order in the potential. This is left for future works.

3.8. Cosmological information content of the convergence PDF

Up to now, this chapter has presented a complete model for the convergence PDF based on the integration along the line of sight of the cumulant generating function of the matter density field itself obtained from first principles with a formalism based on large deviation theory. These developments were both motivated by the goal to push further the use of the large deviation principle in cosmology but also more pragmatically to provide an accurate theoretical description for a promising non-Gaussian observable. Thus the aim of this section is to assess the cosmological information content of the convergence PDF, at first by validating the response of our PDF to cosmological parameter changes again via the use of numerical simulations and then by performing a Fisher forecast based on our theoretical derivatives

3.8.1 Cosmology dependence of the variance and skewness

To get preliminary insights into the cosmology dependence of the PDF derived as above, one can examine the dependence of the cumulants of 2D slices of densities and the lensing convergence on the cosmological parameters. Let me here consider the cold dark matter fraction $\Omega_{\rm cdm}$, the power spectrum amplitude σ_8 and the dark energy equation of state w_0. Specifically, I focus here on understanding the cosmology dependence of the PDF through the effects of the cosmological parameters on its second and third cumulants (the variance and skewness $\langle\kappa^3\rangle_c \equiv \kappa_3$), extending to the dark energy equation of state what was pioneered in Bernardeau et al. (1997). It is clear that the density PDF in the context of LDT is intimately related to both the scale-dependence of the linear variance and the dynamics of the cylindrical collapse. More precisely, different densities ρ scan the linear variance at scales $R\rho^{1/2}$ for a range of values around the radius R. This is encapsulated in the behaviour of the theoretical prediction for the reduced skewness, S_3, of the density at scale R. In an Einstein de Sitter (EdS) universe this quantity is, as already seen, determined by the first logarithmic derivative of the linear variance

$$S_3^{\rm 2D}R) = \frac{\langle\delta^3(R)\rangle}{\langle\delta^2(R)\rangle^2} = \frac{36}{7} + \frac{3}{2}\frac{d\log\sigma_l^2(R)}{d\log R}. \tag{3.55}$$

The first term of this expression comes directly from the spherical collapse dynamics while the second one is a pure effect of the top-hat smoothing of radius R one applies on the field. As such the first term formally depends on the expansion rate of the Universe and therefore on the cosmological parameters but this dependence has been found to be very mild in practice (Bernardeau, 1992; LoVerde, 2014) so that we might keep our implementation of the EdS collapse throughout this analysis. This means that the main dependence on the expansion rate of the Universe will be encoded in the projection and smoothing effects.

Note that $S_3^{\rm 2D}(R)$ does not depend on the overall amplitude of the density fluctuation σ_8 as its dependence cancels out in the logarithmic derivative. It depends however on the scale-dependence of the variance. In figure 3.11, I show how $\Omega_{\rm cdm}$ impacts the linear variance at $z = 0$ and hence the reduced skewness. Also keep in mind that the shape of the full PDF is also sensitive to a combination of the reduced kurtosis S_4 and higher order cumulants, which depend on higher order logarithmic derivatives.

Before the re-scaling by the convergence non-linear variance at the level of the projected CGF, the projection formula allows us to write down the convergence skewness as

$$k_3(\theta, z_s) = \int_0^{z_s} \frac{{\rm d}z}{H(z)}\omega^3(z)S_3^{2D}(\chi(z)\theta)\sigma_{\delta,l}^2(\chi\theta)D^2(z), \tag{3.56}$$

with $D(z)$ the linear growth factor which is the redshift dependence of the linear power spectrum and is given in general by the growing mode of

$$\ddot{D} + 2H(z)\dot{D} - \frac{3}{2}\Omega_{m,0}H_0^2(1+z)^3 D = 0 \tag{3.57}$$

which can be solved analytically for a flat ΛCDM universe as

$$D(z) = \frac{H(z)}{H_0}\int_z^\infty \frac{dz'\,(1+z')}{H^3\,(z')}\left[\int_0^\infty \frac{dz'\,(1+z')}{H^3\,(z')}\right]^{-1}. \tag{3.58}$$

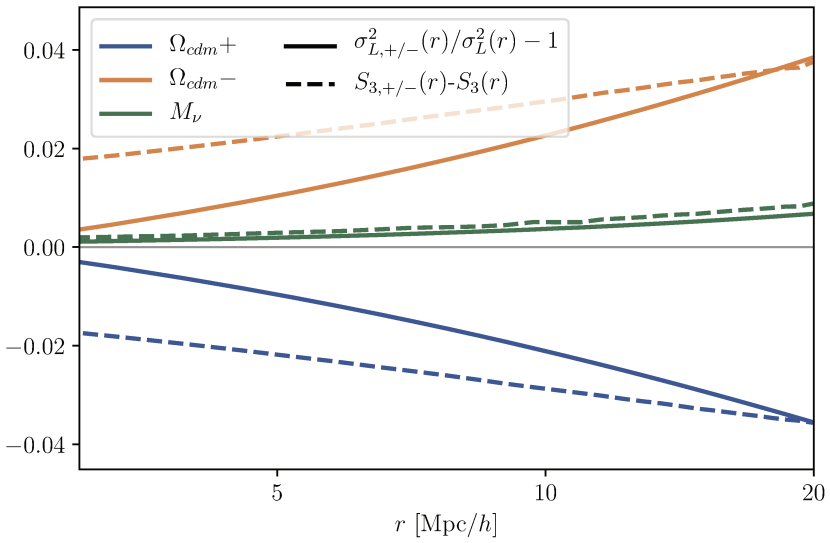


Figure 3.11: Solid lines: The response (residuals as indicated in the legend) of the linear 2D density variance $\sigma^2_{\delta,l}$ at $z = 0$ to different changes in cosmology (denoted by colours) from a fiducial cosmology (see table 3.3). The increments used are those specified in the same table (using the smaller increment for $\Omega_{\rm cdm}$). Dashed lines: The corresponding differences in the reduced skewness S_3^{2D} as calculated using equation (3.55). This figure was published in Boyle et al. (2021).

This thus allows us to give the explicit dependence of k_3 on any cosmological parameter up to the final rescaling by the convergence non-linear variance which amounts to multiplying k_3 by $\sigma^2_{\kappa,nl}/\sigma^2_{\kappa,l}$ and is not expected to impact the general trend of our results. The upper panel of figure 3.12 shows the integrand for k_3 for various cosmologies. The lensing kernel for a source redshift $z_s = 2$ peaks around $z = 0.5$. The middle panel shows the fractional changes in the integrand induced by varying $\Omega_{\rm cdm}$. The plot dissects the different contributions for $\Omega_{\rm cdm}$: the contribution from the $\Omega_{\rm m}$ factor in the lensing weight (sparsely dotted line), which is counteracted by the change in the linear growth factor $D(z)$ (dot-dashed line) and the comoving distance $\chi(z)$ (tightly dotted line). The total change for $\Omega_{\rm cdm}$ depends on the smoothing scale due to the impact of the cylindrical reduced cumulant S_3 and the scale-dependent variance $\sigma^2_{\delta,l}$. The lower panel of figure 3.12 shows the fractional changes in the integrand that result from varying w_0 and the sum of massive neutrinos M_ν. In this case, we see that the change induced by w_0 is independent of the smoothing scale. This is a shared property of all the moments, which allows us to disentangle the $\Omega_{\rm cdm}$ and w_0 contributions from each other even at a single source redshift. Note that while changes in $\Omega_{\rm cdm}$ and w_0 lead to a redshift-dependent change in the integrand of the moments, a change in σ_8 causes a redshift-independent enhancement or diminution. I finally emphasise that while I have used the linear variance here for illustrative purposes, the real model consistently includes the nonlinear variance.

3.8.2 Extension to massive neutrinos

The inclusion of massive neutrino cosmologies in our model for the convergence PDF requires a few adjustments in the implementation of the theory since massive neutrinos affect the PDF

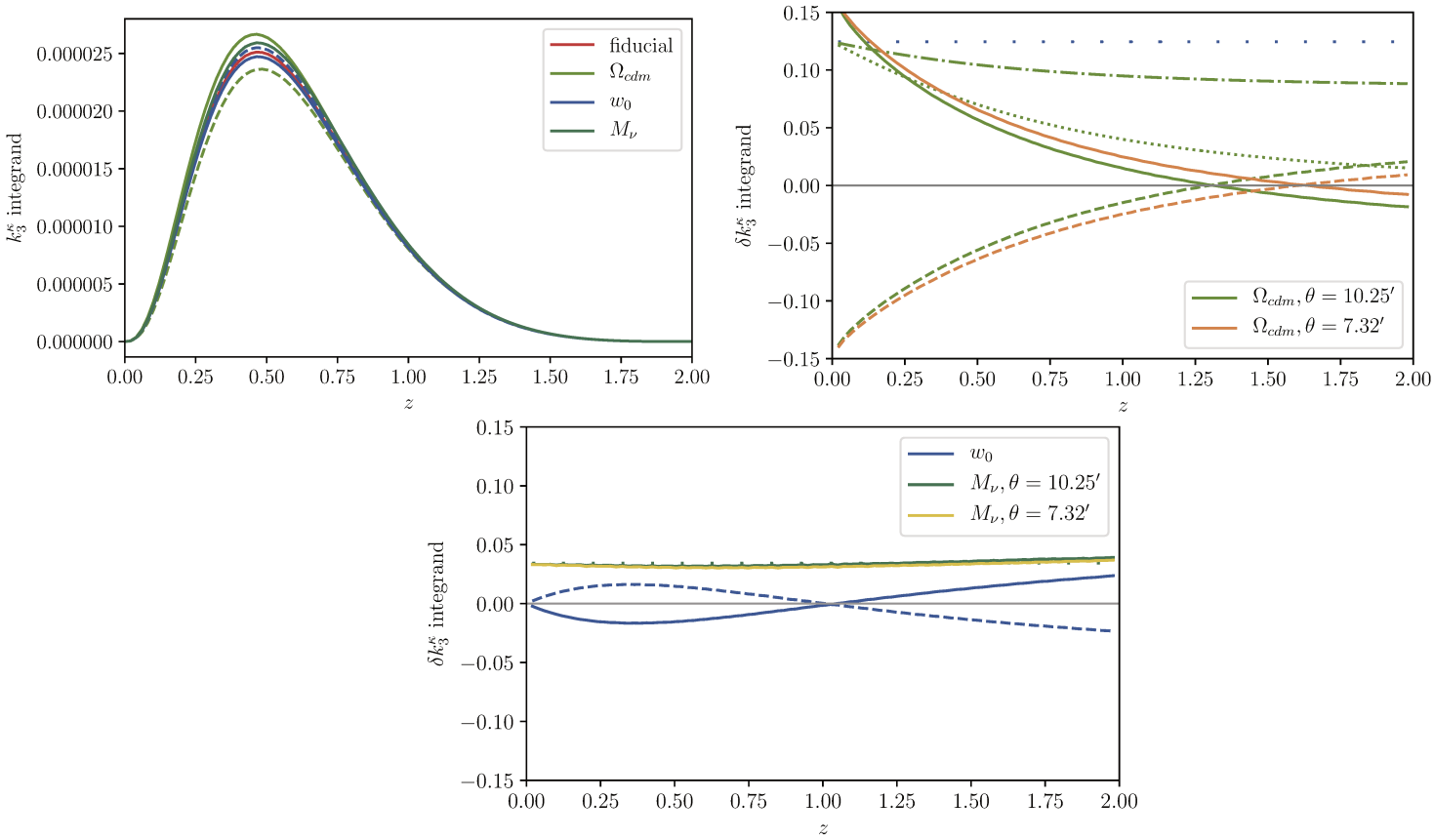


Figure 3.12: Top left panel: The integrand used to calculate the third cumulant of κ for various cosmologies for $\theta = 10$ arcmin. The solid lines correspond to increases in the corresponding cosmological parameters (by the smaller increments in Table 3.3) and the dashed lines to decreases. Top right panel: Fractional changes in the integrand induced by changing $\Omega_{\rm cdm}$ for the two smoothing scales considered for our Fisher forecast. The solid and dashed lines correspond to increases and decreases in the cosmological parameters once again. The additional lines represent the various contributions to the overall differences (see the main text). Bottom panel: Fractional changes induced in the integrand by increasing or decreasing w_0, and by changing M_ν from 0 to 0.15 eV. The changes with w_0 are independent of smoothing scale (so only one set of lines are shown) and the changes with M_ν vary little with smoothing scale (compare the dark green and yellow lines). The M_ν integrand is almost constant with redshift. The dotted dark green line shows the effect of changing only the $\Omega_{\rm m}$ factor in the integrand. This figure was published in Boyle et al. (2021).

in multiple ways. First, the total matter density fraction $\Omega_{\rm m}$ now contains a contribution from Ω_ν. $\Omega_{\rm m}$ enters into the lensing kernel and is by far the dominant effect, as can be seen in figure 3.12 by comparing the solid dark green line (the overall effect of M_ν on the k_3 integrand) and the sparsely dotted dark green line (the effect of changing $\Omega_{\rm m}$ alone). Because the effect of changing $\Omega_{\rm m}$ dominates, the effect of changing M_ν in figure 3.12 is relatively insensitive to redshift and smoothing scale. This creates a natural degeneracy with σ_8.

Next, the scaling of the comoving distance $\chi(z)$ with redshift changes because the Hubble parameter $H(z)$ changes. The scaling of $H(z)$ is now more complicated, as massive neutrinos contribute with a radiative equation of state at early times and gradually transition into forming a contribution like cold dark matter at late times. For an analytic solution to the evolution of the energy density of neutrinos that can be inserted into the Friedmann equation, see Slepian & Portillo (2018). The simplest solution in this case is to use the Boltzmann code CLASS to extract the appropriate evolution of $H(z)$ and $\chi(z)$ with redshift, this is not an issue since we already extract the linear power spectrum using this method.

The unique evolution of massive neutrinos from ultra-relativistic to non-relativistic particles over the course of the history of the Universe has a distinctive effect on the growth of structure. At early times, neutrinos free-stream out of gravitational perturbations and therefore do not contribute to the growth of structure, but they do cluster at late times. This leads to a distinctive scale-dependence in the formation of structures, even on linear scales, with the growth of structure on smaller scales being relatively suppressed. Ultimately, this means that the linear variance $\sigma^2_{\delta,l}(R,z)$ can no longer be evolved with redshift through multiplication by a scale-independent growth factor $D^2(z)$, as the growth factor becomes mildly scale-dependent. For practical implementation, the scale-dependence of the variance must now be independently measured in each redshift slice.

Finally, massive neutrinos make a small contribution to the cylindrical collapse mechanism but LoVerde (2014) showed that for neutrino masses within realistic current bounds, the main effect of massive neutrinos on spherical collapse is to increase the collapse threshold by at most 1%. The impact of such a change can be estimated by varying the parameter ν, and Uhlemann et al. (2019) showed that such a change resulted in a change in the bulk of the 3D matter PDF of less than 1%. For cylindrical collapse, we assume that the effect is also negligible, and find this approximation sufficient to achieve good accuracy.

3.8.3 Validating the PDF derivatives

In this subsection, I compare the predicted response of the PDF to cosmological parameter changes with measurements from the Dustgrain simulation, this response entering the Fisher forecast in terms of derivatives with respect to cosmological parameters. The calculation of these theoretical derivatives requires several ingredients. For each cosmology (see table 3.3), I generate linear and non-linear (Halofit) matter power spectra from CLASS (Blas et al., 2011) at a range of redshifts between $z = 0$ and the chosen source redshift $z_s = 2$. I use these to calculate the linear and nonlinear variance values as a function of redshift, the linear variance is fitted with the analytical form presented earlier to allow the analytical continuation in the complex plane while the non-linear density variance is used to re-scale the density CGF in each slice as opposed to re-scaling the projected CGF by the convergence non-linear variance. CLASS is also used to output $H(z)$ and $R(z)$ for the relevant redshifts. This allows me to

Parameter	Fiducial value	Small Increment	Large Increment
$\Omega_{\rm cdm}$	0.26436	0.0125	0.1
σ_8	0.842	0.034	0.135
w_0	-1	0.04	0.16
M_ν (eV)	0	0.15	-
h	0.6731	0.005	-
n_s	0.9658	0.01	-
$\Omega_{\rm b}$	0.0491	0.0022	-

Table 3.3: The fiducial cosmology and parameter increments chosen for my calculations of the cosmological information content of the convergence PDF, largely chosen for consistency with the Dustrain simulations suite. Those simulations were performed with two step sizes for the three main parameters that were varied ($\Omega_{\rm cdm}$, σ_8 and w_0) as shown. We do not have simulations that vary the parameters below the horizontal line ($h, n_s, \Omega_{\rm b}$), but calculate their effects theoretically using the step sizes shown. This table was published in Boyle et al. (2021).

easily obtain the convergence CGF and then PDF as before by undergoing an inverse Laplace transform.

Contrary to the full-sky simulation by Takahashi, the simulations that we use in this section (because they were run with different cosmologies) have produced very small-area maps (5 deg on a side for Dustgrain). The non-linear density variance with which we rescale our density CGF must take this into account. In these cases, instead of simply rescaling by $\sigma^2_{\rm nl}(\chi(z)\theta)$ at each redshift, the non-linear variance is computed from the integral of the Halofit power spectrum but applying a cut in l (translated to k at each slice) to exclude the large modes beyond the size of the maps with which we compare. This cut is not included when I perform the Fisher forecast.

Simulation-wise, I measure the histogram of the smoothed κ maps in 101 linearly spaced bins for the range $\kappa \in [-0.01, 0.01]$. In principle, one might prefer a finer binning, that for example allows to accurately infer the moments from the binned PDF, however, one should consider that when extracting the Fisher information the Kaufman-Hartlap factor given in equation (3.66) below essentially includes a penalty for too many bins due to the finite number of realisations (for example 256 for Dustgrain or 500 for FLASK) from which the covariance is estimated. Cutting the tails of the PDF also allows to get a more accurate model even though the scale mixing of the lensing kernel is still technically an issue.

In figure 3.13, I compare the derivatives of the PDF with respect to $\Omega_{\rm cdm}$, σ_8 and w_0 as obtained using the theory and the Dustgrain simulation at two relevant scales. I also show a theoretical derivative for M_ν, although we cannot directly validate it with Dustgrain (we validate it by comparison with another simulation – MassiveNuS – instead in figure 3.14). We find the results to be in very good agreement. We note that with the number of realisations we have, the smaller step sizes provided for the simulations (see table 3.3) result in significant numerical noise (see the dotted lines in figure 3.13). This kind of noise can result in exaggerated constraints by artificially breaking degeneracies between parameters. The larger step sizes provide much smoother derivatives, but we find that even those contain some small amount of noise that can artificially enhance their constraining power by a significant amount when inserted into a Fisher matrix. This presents an obvious advantage

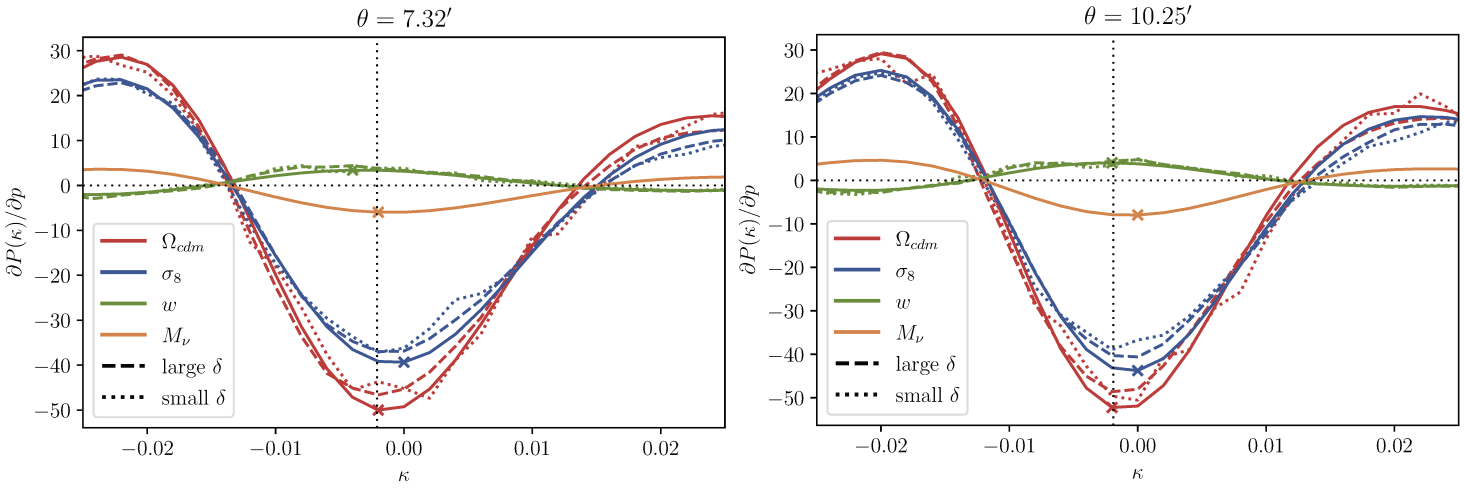


Figure 3.13: Derivatives of the lensing convergence PDF with respect to $\Omega_{\rm cdm}$, σ_8, w_0 and M_ν computed for large/small steps in the simulation (dashed/dotted) and the theory for large steps (solid) at two smoothing scales, $\theta = 7.32'$ and $\theta = 10.25'$. In the M_ν case, only a theoretical line is shown, and this is validated using the MassiveNuS simulations instead in figure 3.14. One can see that the small parameter steps in the simulation suffer from numerical noise, which is why I rely on the large steps to validate our theoretical predictions. When calculating the theory derivatives for validation purposes, I include a large-scale cut in l when calculating the variance at $l = 36$, corresponding to the patch size of Dustgrain. The vertical dotted lines correspond to the locations of the peaks in the fiducial PDFs. The coloured crosses indicate the extrema of the derivatives. This figure was published in Boyle et al. (2021).

to having a theory with which to model observables in such calculations. In addition, our validated complete cosmological model for the PDF allows us to extend our forecasts to parameters beyond those varied in particular simulations. Note that while the general shape of derivatives might look similar at a first glance, there are two hints for potential degeneracy breaking. If derivatives have either different zero crossings or different extrema locations, this means they are not proportional to each other. The first can be seen for example in the positive κ region for the red $\Omega_{\rm cdm}$ and the blue σ_8 lines. The second can be observed from the green line for w_0 whose peak is at a different location from the rest of the dips.

In figure 3.14, I use the MassiveNuS simulations as a basis for determining how well our theory captures the effects of massive neutrinos on the PDF. The fiducial MassiveNuS cosmology (referred to as model 1b in Liu et al., 2018) has cosmological parameters $\Omega_{\rm m} = 0.3$, $\sigma_8 = 0.8295$ and $M_\nu = 0.1$ eV. This is used with a model with the same $\Omega_{\rm m}$ but without massive neutrinos, which has $\sigma_8 = 0.8523$ (model 1a) to calculate the difference shown in figure 3.14. All relevant cosmological parameters are provided in Table 1 of Liu et al. (2018) and its caption. We see from figure 3.14 that the effects of adding massive neutrinos to the cosmology are well captured by the theory.

3.8.4 Fisher forecasts

Let me here briefly recap the Fisher forecast formalism. The basic building block of contemporary analyses is the *likelihood* function which is defined as the probability that an experiment yields the observed data given a theory, for example a set of cosmological parameters $\boldsymbol{\theta}$

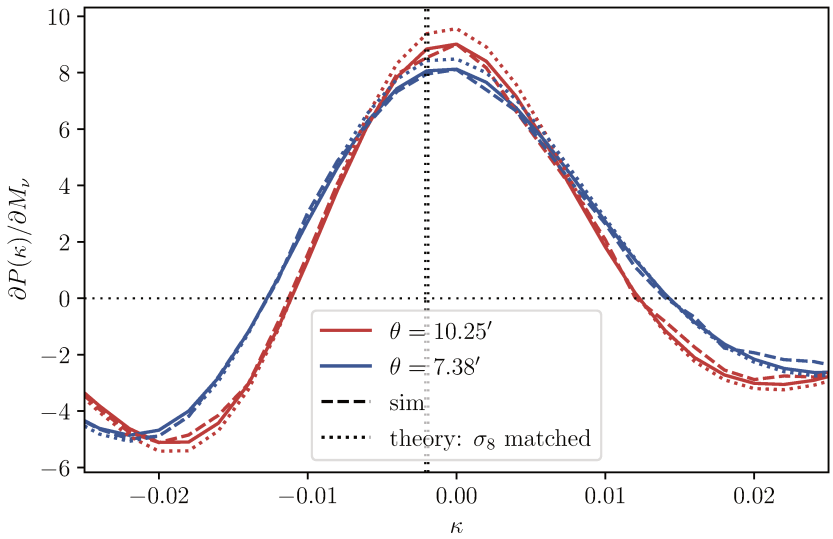


Figure 3.14: Derivative of the lensing convergence PDF with respect to M_ν computed as a single-sided derivative using the difference between the fiducial cosmology from MassiveNuS (which has a total neutrino mass M_ν of 0.1 eV) and their corresponding cosmology without massive neutrinos. This particular comparison is shown only for validation purposes. I calculate the actual theoretical derivatives used in the Fisher matrix calculations for the cosmology given in Table 3.3. Here we include a cut in l at $l = 60$, corresponding to the MassiveNuS patch sizes. The smaller smoothing scale here is set to $7.38'$ instead of $7.34'$ simply because it corresponds to a whole number of pixels for the MassiveNuS resolution. The direction of the derivative is reversed compared to that in Figure 3.13 because σ_8 is not fixed here and adding massive neutrinos lowers σ_8. The dotted 'σ_8 matched' lines show how the derivative looks if the PDF with massive neutrinos in the theoretical derivative is replaced by a PDF for a cosmology without massive neutrinos but with equal σ_8 ($\sigma_8 = 0.8295$ from Liu et al., 2018). The difference between the solid and dotted lines therefore corresponds to the orange line in Figure 3.13. The vertical dotted lines correspond to the locations of the peaks in the fiducial PDFs (the leftmost is for the smaller scale). This figure was published in Boyle et al. (2021).

$$\mathcal{L}\left(\text{data} \mid \boldsymbol{\theta}\right) \equiv P\left(\text{data} \mid \boldsymbol{\theta}\right). \tag{3.59}$$

This function then has to be estimated theoretically, for example assuming it is Gaussian and demonstrate numerically that it is indeed close to be the case as I do for the convergence PDF, or using some other theoretical techniques to compute it. Estimating numerically the likelihood for some experiment in our context can for example thus account to simulate the same object (here the PDF) in the same conditions (here the cosmological model) multiple times and draw its distribution around the most probable value, in a large sense, then assumed to be the "true" value.

However, real parameter estimation does not stop there since we would in practice rather like to compute $P\left(\boldsymbol{\theta} \mid \text{data}\right)$ which can be obtained by means of *Bayes' theorem*

$$P\left(\boldsymbol{\theta} \mid \text{data}\right) = \frac{P\left(\text{data} \mid \boldsymbol{\theta}\right) P\left(\boldsymbol{\theta}\right)}{P\left(\text{data}\right)}. \tag{3.60}$$

Since the denominator (called the *evidence*) is just a constant normalisation that does not depend on the cosmological parameters then it will not affect the most probable values for the parameters and their error bars and we can thus usually ignore it. The left hand side is called the *posterior* and is thus the product of the likelihood and the *prior* which contains as the name implies prior information about the parameters. In general one has to be very careful about priors since a flat one does not exactly imply no information at all on the parameters. For example a flat prior on a parameter α – that is assuming that intervals of α equally likely – is not equivalent to a flat prior on α^2... In the context of cosmological parameters estimation, the priors can be the previous findings of an earlier experiment. Nevertheless, the game now amounts to finding the parameter space where the posterior is the largest, for a flat prior this for example accounts to finding the maxima of the likelihood by differentiation and the $1-\sigma$ error bar (or 68% confidence level) is obtained from the shape of the likelihood around those maxima.

As theorists however we do not have any data but rather predict new signal, for example the PDF of the convergence field, and we thus want to answer a different question: how well do we expect a given experiment to determine cosmological parameters? This is called forecasting. In principle this also requires knowledge about the full shape of the likelihood but if it is close to Gaussian, then a Taylor expansion around the true values of the parameters makes its curvature (its second derivatives taken at the true values of the parameters) the dominant term. We then call it the Fisher matrix F_{ij} which thus quantifies the amount of information that a given experiment can provide about a set of parameters. In practice we also rather define it as the second derivatives of the log of the likelihood function, where the log is taken to conveniently transform multiplication of probabilities into sums

$$F_{ij} \equiv -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial\theta_i \partial\theta_j} \right\rangle \Bigg|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}}, \tag{3.61}$$

where $\bar{\boldsymbol{\theta}}$ are the true values of the parameters.

As a consequence, assuming a Gaussian likelihood and for a (combination of) statistics $\boldsymbol{s}$, the Fisher matrix is now defined as

$$F_{ij} = \sum_{\alpha,\beta} \frac{\partial s_\alpha}{\partial \theta_i} C^{-1}_{\alpha\beta} \frac{\partial s_\beta}{\partial \theta_j}, \tag{3.62}$$

where s_i is element i of the statistic $\boldsymbol{s}$ and $C_{\alpha,\beta}$ are the elements of the covariance matrix of measurements $\hat{\boldsymbol{s}}$ of $\boldsymbol{s}$, defined as

$$C_{\alpha\beta} = \langle (\hat{s}_\alpha - \langle \hat{s}_\alpha \rangle)(\hat{s}_\beta - \langle \hat{s}_\alpha \rangle) \rangle. \tag{3.63}$$

The Fisher matrix then allows us to determine the error contours on a set of cosmological parameters under this Gaussian likelihood approximation. The inverse of the Fisher matrix gives the parameter covariance. The error on the parameter θ_i, marginalised over all other parameters, is given by

$$\delta\theta_i \geq \sqrt{(F^{-1})_{ii}}. \tag{3.64}$$

A Fisher analysis has a number of limitations: as said it only yields realistic error bars if measurements of the considered data vectors have Gaussian noise and if the responses of these data vectors to changing cosmological parameters are close to linear. I thus demonstrate that the distribution of individual bins of the PDFs measured in the simulations are sufficiently close to a Gaussian distribution in figure 3.15. I there focus on the lowest and highest two κ bins I consider for the Fisher analysis (see the following section) and make a histogram of the fluctuations of the individual realisations around the mean in units of the measured variance. From this figure, we can expect a small impact of residual non-Gaussianity on the total width of our forecasted contours. Note that, in contrast, the moments of the lensing convergence can have strongly non-Gaussian distributions around the mean, in particular in small simulation patches as demonstrated in the discussion of moments-based constraints in appendix A.4 and are not well-suited for a standard Fisher analysis. Fully realistic data analyses might have to account for systematic effects by marginalising over additional nuisance parameters. However, the main focus of this study is to explore the complementarity between the two-point correlation function and the one-point lensing convergence PDF as cosmological probes. As a result, I expect these limitations to have limited impact on our findings.

3.8.5 Covariance matrices

In the following, the data vector $\boldsymbol{s}$ entering equation (3.62) will either consist of histogram bins of measurements of the convergence PDF, of measurements of the two-point correlation function of convergence in a set of angular bins or of a combination of both. In the cases where $\boldsymbol{s}$ includes PDF measurements we estimate the covariance matrix that enters equation (3.62) from a set of $N_{\rm sim}$ simulated measurements $\hat{\boldsymbol{s}}_i$, $i = 1, \ldots, N_{\rm sim}$, as

$$\hat{C}_{\alpha,\beta} = \frac{1}{N_{\rm sim} - 1} \sum_i (\hat{s}_{i,\alpha} - \bar{s}_\alpha)(\hat{s}_{i,\beta} - \bar{s}_\beta) \ , \tag{3.65}$$

where $\bar{\boldsymbol{s}} = \sum_i \hat{\boldsymbol{s}}_i / N_{\rm sim}$ is the mean of all measurements. To compute the Fisher matrix with equation (3.62), we need to estimate the inverse covariance matrix or precision matrix. Since

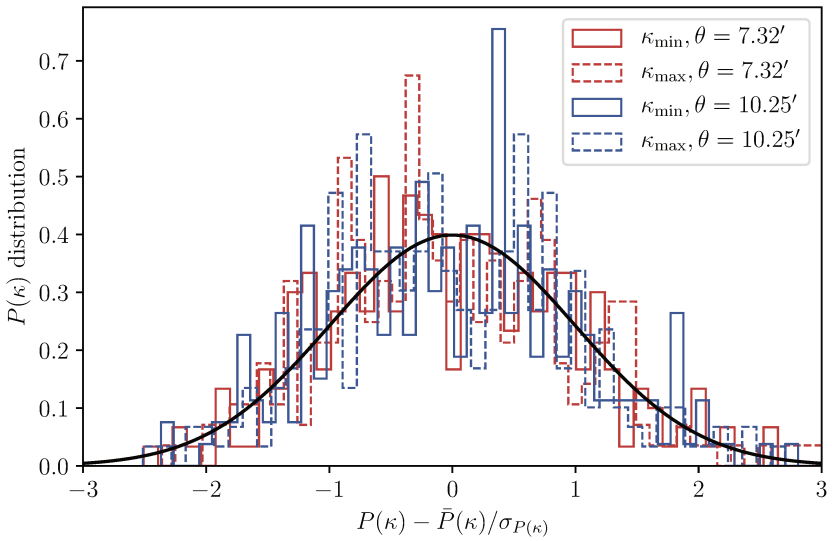


Figure 3.15: The distribution of realisations around the mean PDF in the lowest (red) and highest (blue) bins for the smaller smoothing scale (solid) and the larger scale (dashed). The black line shows a Gaussian of zero mean and unit variance, which is the expectation for Gaussian noise and a good approximation to all distributions. This figure was published in Boyle et al. (2021).

matrix inversion is a non-linear operation, the noise in the above estimate of the covariance elements $\hat{C}_{\alpha,\beta}$ will lead to a bias in the elements of the precision matrix. Through a miracle of nature this bias is just a factor multiplying the entire matrix, the so called Kaufman-Hartlap factor (Hartlap et al., 2006),

$$h = (N_{\rm sim} - 2 - N_s)/(N_{\rm sim} - 1), \tag{3.66}$$

where N_s is the number of data points in $\boldsymbol{s}$. Please note that the Kaufman-Hartlap factor (but also the more advanced treatment of Sellentin & Heavens, 2016) only corrects the width of parameter contours for bias from covariance estimation noise. This is sufficient for a Fisher analysis like the one presented here, but it is not sufficient for parameter estimation based on an actual measurement of $\boldsymbol{s}$. In the latter case, covariance noise will not only impact the width of parameter contours but also introduce additional scatter to the contour location. This needs to be taken into account independently of the Kaufman-Hartlap factor, as *e.g.* described in Dodelson & Schneider (2013) or even more completely in Percival et al. (2014) (see also figure 1 of Friedrich & Eifler, 2018, for a visualisation of this effect).

In figure 3.16 we show the correlation matrix

$$\mathrm{Corr}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, \tag{3.67}$$

for measurements of the convergence PDF at two smoothing scales. This matrix visualises the amount of correlation throughout the data vector elements. I compare the correlation matrices obtained from 256 maps from the Dustgrain simulations and from 500 log-normal

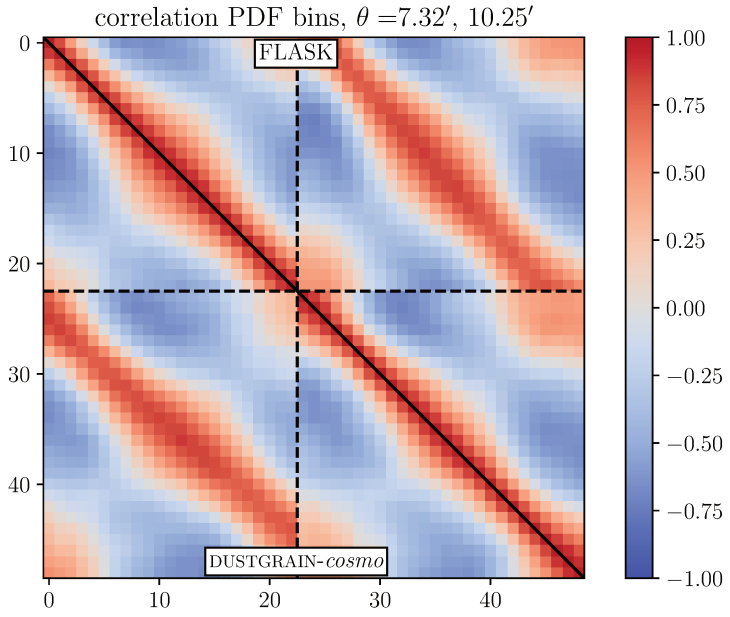


Figure 3.16: The reduced cross-correlation matrix (equation (3.67)) of the weak lensing convergence PDF at radii $\theta_1 = 7.32'$ and $\theta_2 = 10.25'$ and source redshift $z_s = 2$ without shape noise. I display bins [-0.024,0.026] for θ_1 and [-0.022,0.022] for θ_2, which corresponds to cutting $[2\%, 5\%]$ probability in low/high κ. The black line indicates the diagonal. The lower triangle shows the result from the 256 κ maps created from one Dustgrain N-body simulation with their randomisation procedure and the upper triangle shows the result from 500 independent full-sky maps created from FLASK. This figure was published in Boyle et al. (2021).

maps* generated using the publicly available FLASK tool (Xavier et al., 2016). In practice, I implement the latter in our Fisher matrix calculations, but we see that both methods are in reasonable agreement. Naturally, in both cases, the measured covariances are re-scaled to the area of a Euclid-sized survey, assuming that the covariance is proportional to survey area. In the two blocks on the diagonal in figure 3.16 we see that, as expected, neighbouring PDF bins are positively correlated, while intermediate underdense and overdense bins are anti-correlated with each other. Note that the tails of the PDF, which are excluded in the plot and our analysis, are strongly correlated with each other and anti-correlated with the peak. Additionally, the PDFs at different scales are strongly but not perfectly correlated with each other. This is expected, as the two smoothing apertures partially (but not perfectly) overlap. The cross-correlations between PDF bins of different scales look very similar to the bin correlations for the individual PDFs, because the matter clustering changes mildly with radius (see Figure 5 in Uhlemann et al., 2017a).

In all contours that only consider the convergence two-point function I use an analytic covariance model in order to circumvent noise (and potential small-scale numerical artefacts) associated with covariance estimation from FLASK. The covariance model is then based on

*I have already mentioned that the convergence field is in practice found to be close to a log-normal field with the variance and skewness obtained from high-order perturbation theory or numerical simulations. The basic idea of FLASK is thus from a convergence power spectrum and a given skewness to quickly generate convergence maps. This is very useful for example to estimate covariance matrices as I do in this thesis.

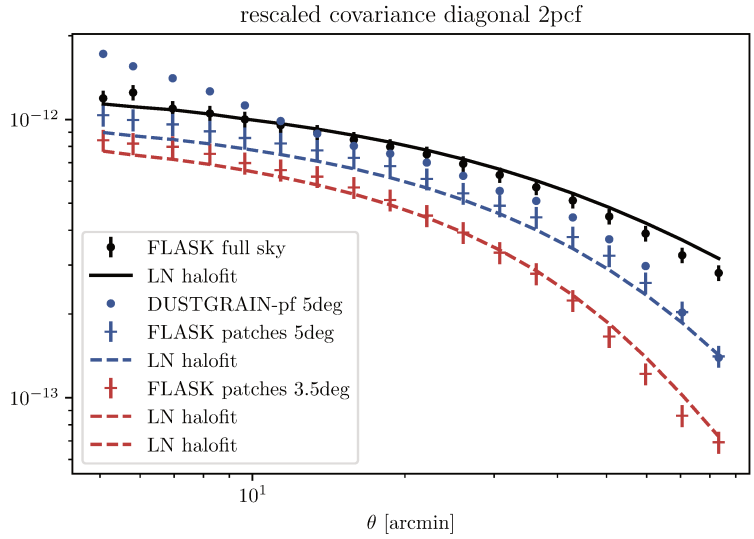


Figure 3.17: Variance of the κ two-point correlation function bins rescaled to a common sky area of 15000 deg^2. The theoretical prediction from the log-normal model (black line) is validated against the measurements from full sky FLASK maps (black data points). We also show that considering small patches with mean subtraction (blue and red crosses) underestimate the covariance and are equivalent to imposing an $l_{\rm min}$ cut appropriate for the different patch sizes in the theory (dashed lines). Those predictions are in broad agreement with the measurements from small patch simulations of Dustgrain (blue points), with deviations driven by residuals of the two-point correlation with respect to Halofit (see figure B2 of Boyle et al. (2021)). This figure was published in Boyle et al. (2021).

the log-normal covariance model by Hilbert et al. (2011) but uses the procedure of Friedrich et al. (2018) to fix the parameters of the log-normal distribution as well as the analytic treatment of sky curvature and bin averaging that is detailed in Friedrich et al. (2020). I do not further detail this model since I am more focused on the work performed on the PDF. Nevertheless I find that this analytic covariance agrees well with the corresponding estimate from FLASK for the range of scales (bigger than a few arcmin) for which FLASK can be reasonably applied, see figure 3.17.

3.8.6 Forecasted Constraints for $\Omega_{\rm m}$, σ_8, w_0 and M_ν

In this subsection, I finally quantify the information content of the lensing convergence PDF on some key wCDM cosmological parameters using the Fisher matrix formalism. After having validated our theoretical predictions for the PDF and the Halofit two-point correlation function against simulation measurements, we perform our forecast using purely those inputs. This is important because, as previously mentioned, apparently negligible amounts of numerical noise in the PDF derivatives extracted from simulations can lead to artificially strong constraints. Note that though I did not consider it to validate the PDF derivatives, I know include shape noise in the cosmological constraints estimations. For the simulated maps, I include shape noise by adding a white noise Gaussian random map to the 'raw' simulated convergence map before smoothing, keeping the seed fixed for a given realisation when varying cosmologies. For the theoretical predictions, I convolve the theoretical 'raw' convergence PDF at the desired smoothing scale with a Gaussian of the appropriate width.

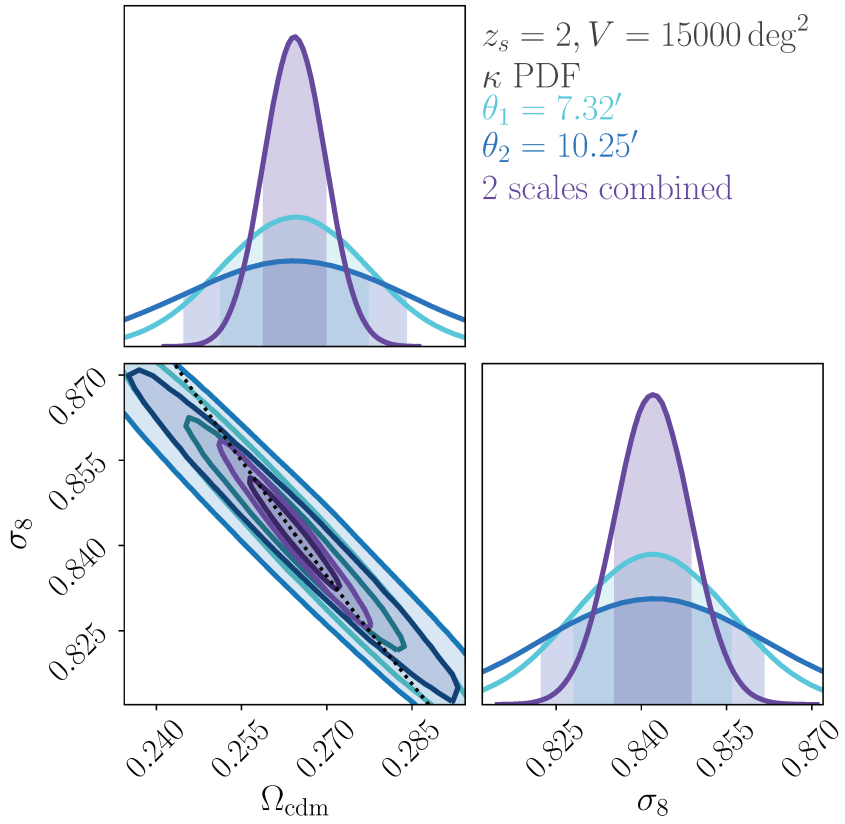


Figure 3.18: Fisher forecast constraints on Ω_{cdm} and σ_8 for a Euclid-like survey from the weak lensing convergence PDF at $z_s = 2$. Contours are shown for the PDF measured for a single smoothing scale of $\theta_1 = 7.32'$ (cyan), a single scale of $\theta_2 = 10.25'$ (blue), and the two scales combined (purple). It is clear that the use of two smoothing scales significantly improves the constraints. The combined PDF contour is oriented along the line of constant $\Sigma_8 = \sigma_8\sqrt{\Omega_{\mathrm{m}}}$ (black dotted line). This figure was published in Boyle et al. (2021).

The number density of observed galaxies is set to 30 arcmin^{-2} which is a bit unrealistic but works as if all the observed galaxies of the Euclid mission contributed to the signal.

Let us now analyse the cosmology-constraining power of the weak lensing convergence PDF and its complementarity to the two-point correlation function. Figure 3.18 shows constraints from the convergence PDF for two individual smoothing scales and their combination. It is clear that combining two smoothing scales helps reduce the degeneracy between Ω_{cdm} and σ_8 because the Ω_{cdm} derivative is a function of the smoothing scale (see figure 3.12). Notice that the degeneracy direction aligns roughly with the $\Sigma_8 = \sigma_8\sqrt{\Omega_{\mathrm{m}}}$ direction.

Figure 3.19 compares the PDF constraints on Ω_{cdm} and σ_8 at two scales to other probes: the two-point correlation function with $\theta_{min} = 5'$ and CMB-based constraints from Planck 2018 (Planck Collaboration, 2020). The Planck constraints are extracted from the `TTTEEE\_lowl\_lowE` chains in the Planck legacy archive for the `base` cosmology. This corresponds to data from the Planck TT, TE and EE spectra at $l > 30$ plus low-l TT and EE spectra. In later figures, I use instead the `base\_w` dataset for the wCDM cosmology or `base\_mnu` for the ΛCDM+M_ν cosmology. The covariance matrices are generated specifically using $\Omega_{\mathrm{cdm}}/h^2$

as the $\Omega_{\rm cdm}$ value for each run (i.e. not using the values provided for $\Omega_{\rm m}$). It is important to note that all the Planck constraints here come marginalised over the standard ΛCDM parameters as a minimum (including h, n_s, $\Omega_{\rm b}$) while our PDF and two-point correlation function Fisher matrices are marginalised only over the parameters shown in each contour plot.

We see that the constraints in Figure 3.19 compare favourably with those from the two-point correlation function, although the degeneracy direction in both cases is similar. This is due to the fact that we have fixed all other cosmological parameters here. The degeneracy directions can differ as soon as one varies more parameters. I demonstrate this effect in figure 3.20, which extends the calculation to include constraints on w_0 and looks more promising. Here adding information from the PDF has the potential to break degeneracies between w_0 and both $\Omega_{\rm cdm}$ and σ_8, potentially tightening constraints considerably. We see a similar behaviour in figure 3.21, which shows simultaneous constraints on $\Omega_{\rm cdm}$, σ_8 and M_ν. It is not possible to extract meaningful constraints on M_ν using either CMB or large-scale structure information alone, but the PDF information content compares well with that from the two-point correlation function, and the degeneracy directions are shifted somewhat, allowing for potential degeneracy breaking by combining the two probes.

The Fisher matrix forecasting method can also be used to analyse the robustness of these results. For instance, implementing the full differential equations for cylindrical collapse instead of our EdS parametrisation makes no difference to the results.

3.8.7 Extended Parameter Set Constraints

The PDF derivatives for the parameters presented in figures 3.18 to 3.21 have all been validated using simulations. It is thus reasonable to assume that it is then safe to extend the application of the model to other cosmological parameters to produce more quantitative results relevant for a realistic future surveys. While an in-depth analysis of the relevant degeneracies and a realistic redshift distribution is beyond the scope of the work presented here, I analyse the key effects of varying the full set of cosmological parameters. In Table 3.4, I show forecasted cosmological constraints from the PDF marginalised over the standard ΛCDM parameters, and combined with the CMB prior from Planck described in the previous subsection.

As we would aim for these results to represent more realistic constraints, I also choose to be more conservative with our shape noise parameters, adopting $n_{\rm gal}$ of 8 arcmin^{-2} instead of 30 arcmin^{-2}, which approximately doubles the shape noise standard deviation but is closer to the value we would have combining two bins of a Euclid-like experiment (that is moving from a 10 to 5 bins strategy). I do this because obviously in practice the total number of galaxies would not be located at a single source redshift, but instead distributed across different redshift bins according to the source galaxy distribution $n_s(z_s)$ as already mentionned several times over the course of this thesis.

Overall, the constraints in table 3.4 are very promising. The constraints in the ΛCDM case are comparable, but the PDF significantly outperforms the two-point correlation function once a free neutrino mass or dark energy equation of state is introduced. We generally find that the resulting contours have quite similar (though not precisely identical) degeneracy directions, as was also seen by Patton et al. (2017), for example. In the ΛCDM case, strong

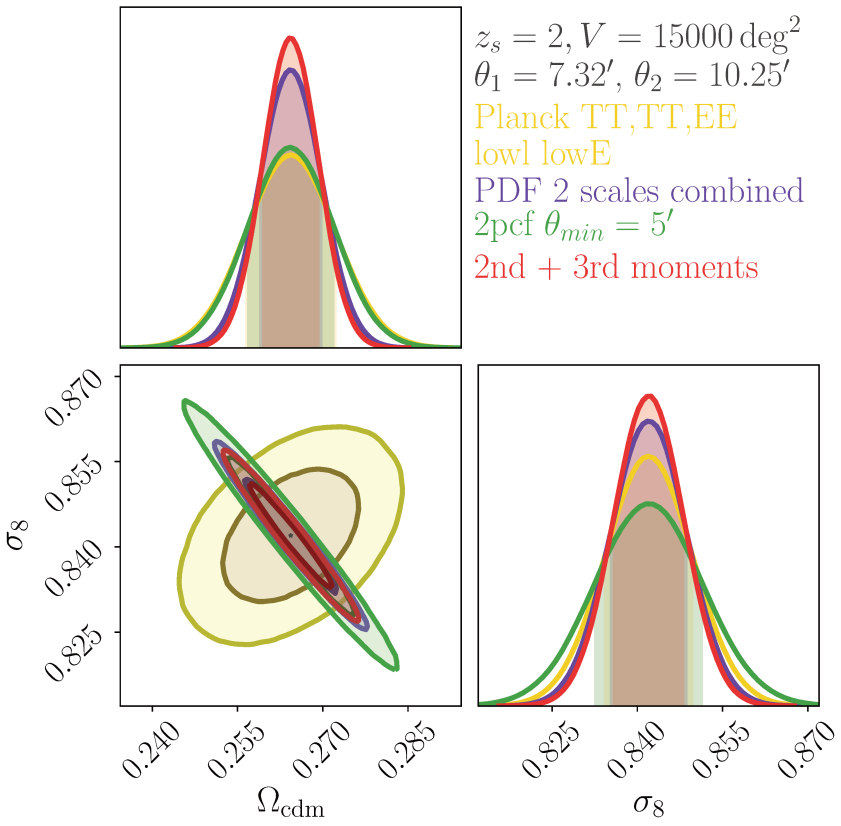


Figure 3.19: Same as figure 3.18, but presenting a comparison of the constraints from the PDF with two smoothing scales (purple) to constraints from other probes. Contours for the two-point correlation function with $\theta_{min} = 5'$ are shown in green. CMB-based constraints from the Planck 2018 TT,EE,EE+lowl+lowE dataset are shown in yellow (see text for details) and show a clearly complementary degeneracy direction. Constraints from the second and third moments alone (pink) slightly outperform the PDF. We note this is not a contradiction as we use a truncated PDF but that this should be taken with some caution. Some details on the use of moments for constraints are given in appendix A.4. This figure was published in Boyle et al. (2021).

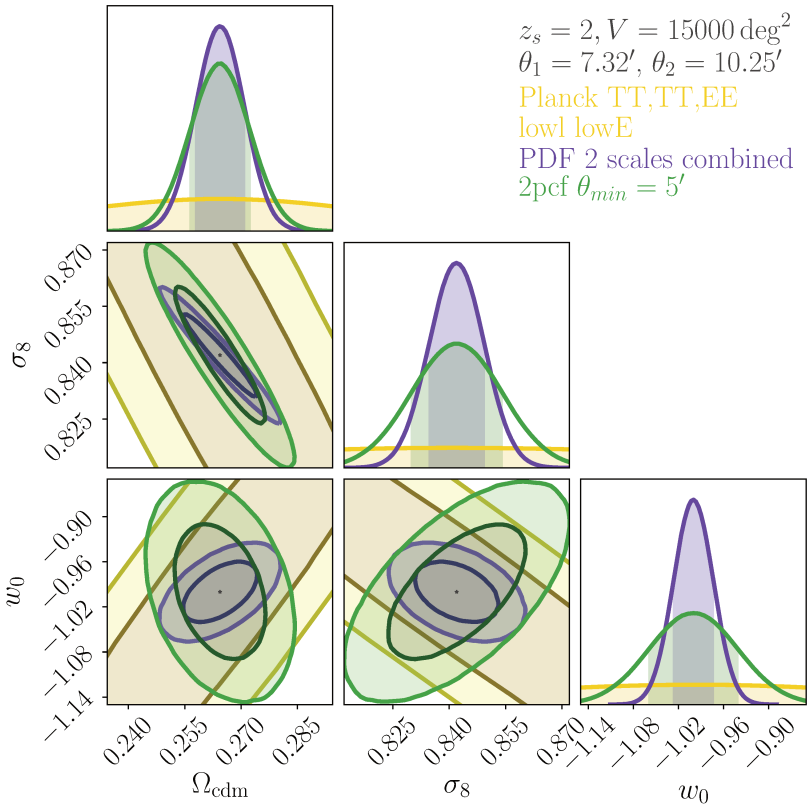


Figure 3.20: Same as figure 3.19, but extended to include the dark energy equation of state w_0 as an additional free parameter. This figure was published in Boyle et al. (2021).

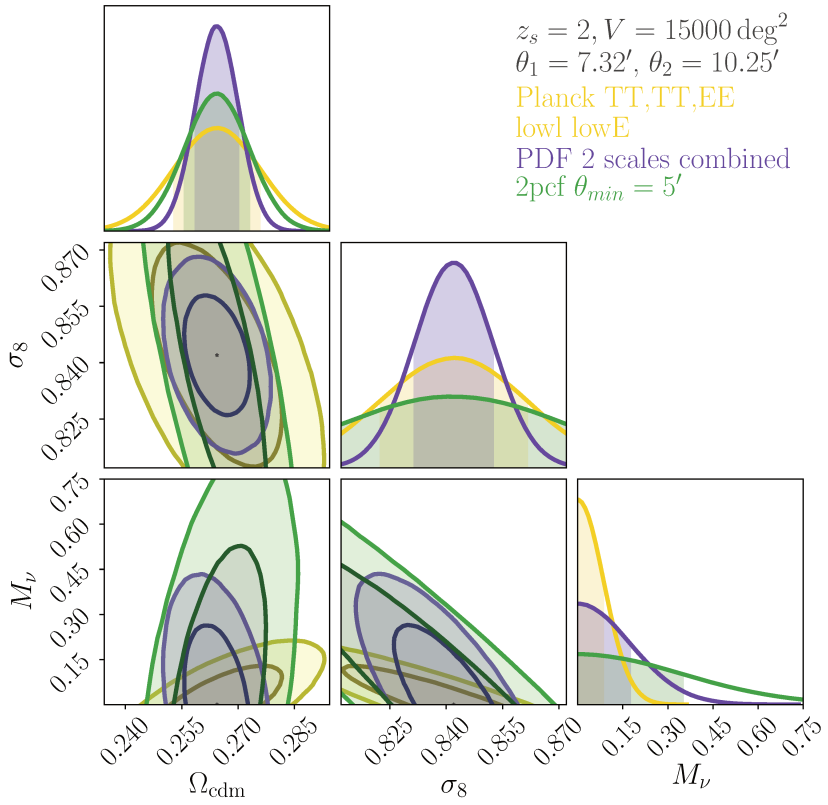


Figure 3.21: Same as Figure 3.19, but extended to include the total mass of neutrinos M_ν as an additional free parameter. This figure was published in Boyle et al. (2021).

Planck+PDF vs. Planck	ΛCDM	ΛCDM+M_ν	wCDM
$\Omega_{\rm cdm}$	+55%	+43%	+65%
σ_8	+43%	+40%	79%
$\Omega_{\rm b}$	+52%	+37%	+67%
n_s	+25%	+24%	+20%
h	+53%	+39%	+74%
M_ν (eV)	-	+31%	-
w_0	-	-	+78%
Planck+PDF vs. Planck+2pcf	ΛCDM	ΛCDM+M_ν	wCDM
$\Omega_{\rm cdm}$	0%	+23%	+36%
σ_8	+8%	+32%	+35%
$\Omega_{\rm b}$	+0%	+26%	+37%
n_s	+0%	+2%	+4%
h	+0%	+25%	+36%
M_ν (eV)	-	+27%	-
w_0	-	-	+40%

Table 3.4: (Upper panel) Percentage improvements in the 1-σ constraints from Planck (`TTTEEE_lowl_lowE`) when adding the convergence PDF at two combined smoothing scales at $z_s = 2$ from a Euclid-like survey. (Lower panel) Percentage improvements in the constraints from Planck combined with the PDF over constraints from Planck and the two-point correlation function. Note that we take a more conservative approach to shape noise here compared to in figures 3.18-3.21 (see text). This table was published in Boyle et al. (2021).

constraints from Planck mean the advantage of the PDF over the two-point correlation function seen in Figure 3.19 becomes less relevant. When additional free parameters that are poorly constrained by the Planck covariance matrix are introduced, this is no longer the case. The PDF can constrain M_ν over 25% better and w_0 about 40% better than the two-point correlation function. An examination of the individual Fisher matrices reveals that the advantage of the PDF lies primarily in its ability to better constrain σ_8 than the two-point correlation function. Of course, both M_ν and w_0 are anti-correlated with σ_8, resulting in the improved constraints we see in table 3.4.

Finally, note that combining Planck with both the PDF and two-point correlation one could expect a modest improvement of 15% on σ_8 constraints for base LCDM. For the extended models we see up to a 50% improvement for most parameters in the wCDM and νLCDM cosmologies, except for n_s, which is still strongly constrained by Planck. In particular we notice a 40-45% improvement on σ_8, M_ν, w_0 and h. Note that the improvement of the PDF over the two-point correlation tends to be reduced by the presence of shape noise. We expect that when repeating a similar analysis at lower source redshifts $z_s \sim 1$ with a larger smoothing scale ~ 20 arcmin (thus keeping the degree of non-linearity roughly the same as for our current setup), the impact of shape noise will be smaller and the PDF potentially even more powerful.

3.9. BNT convergence

Throughout this chapter, I have established in some details a model for the convergence CGF and PDF based on the integration along the line of sight of the one-point statistics of the matter density field. I have presented its validity domain that is somehow reduced because of the scale mixing from the lensing kernel and even presented a forecast of the cosmological information content this formalism can aim at probing. There thus only remains one final touch that I would like to add to this chapter. Indeed I have earlier presented the BNT transform that boils down to narrowing the lensing kernel to an engineered set of physical scales and redshifts thus allowing to separate the contributions to lensing of different regimes of the non-linear formation of structures. From our theorist point of view and in the context of our model for the convergence CGF and PDF this is very exciting since it allows us to apply our formalism to sources at smaller redshift and for smaller angular scales. More precisely, the range of scales and redshifts of the matter density field accurately described by the spherical collapse dynamics does not change, only the weight of their contribution to the new BNT lensing observables.

Nothing changes in the previously presented formalism and its implementation, only for the fact that the lensing kernel is now replaced by its nulled counterpart. To compare the BNT convergence PDF to the Takahashi simulation (we do not need Dustgrain anymore since we do no care about having only one cosmology), I select 8 redshifts from $z_s = 0.57$ to 1.6 so as to roughly mimic the tomographic strategy of the Euclid photometric survey but limited by the available redshifts in the simulation. I then apply the three-plane solution to every set of three neighbouring planes (my choice of normalisation is always $p_3 = 1$) and obtain the effective lensing kernels roughly similar to figure 1.12.

At the level of the simulation and for every set of 3 simulated κ-maps at the chosen redshifts, I take their linear combination with appropriate weights and filter the obtained nulled convergence maps with a top-hat window function of angular radius $\theta = 10$ arcmin. Similarly to the beginning of this chapter, I then compute the theoretical CGFs for the nulled convergences as well as measure it in the maps. Once again the Gaussian contribution is removed to focus on the higher-order contributions and an excellent agreement is found between the simulated and theoretical CGFs in figure 3.22, which is to be contrasted with the case before nulling shown in figure 3.1. As expected, the nulling strategy has allowed us to remove the non-linear effect of the smaller scales and recover a situation where our approach based on cylindrical collapse is accurate enough.

Table 3.5 shows the fitted convergence variances and figure 3.23 displays the resulting PDFs of the 6 nulled convergence maps for a fixed opening angle of $\theta = 10$ arcmin. They are, as for the CGFs, found to match remarkably well with the theoretical predictions: peaks are described within percent accuracy and tails are accurately captured within a 2-3 σ range depending on the redshifts considered. Also note that the inaccuracy that seems to appear at 2 σ in the low-κ regime for the red and orange curves happens at a very low probability not shown in the log panel since the PDF is not symmetric. By essence, the BNT transform considerably reduces scale mixing and will thus be more likely to significantly enhance the accuracy of the theoretical predictions. Note that at these scales and redshifts, the PDFs are all highly non-Gaussian and the convergence variance, nulled or not, is not a good probe for the level of non-linearity as opposed to the variance of the density field. It is even more the

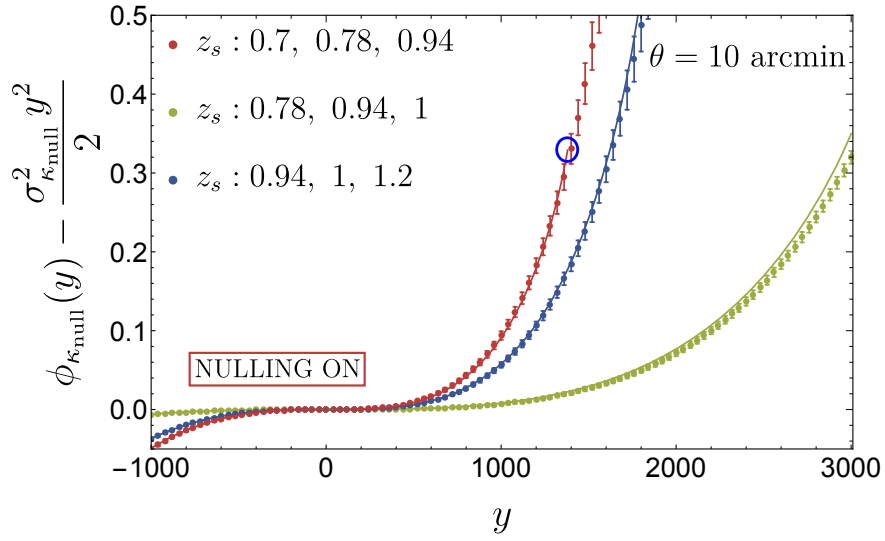


Figure 3.22: CGF of the BNT convergence for different source redshifts and opening angle $\theta = 10$ arcmin. Data points are taken from the simulation after implementation of the nulling procedure. Blue circles indicate critical points as already discussed for figure 3.1. This figure was published in **Barthelemy** et al. (2020a).

$z_{s,1}$	0.57	0.71	0.78	0.94	1	1.2
$z_{s,2}$	0.71	0.78	0.94	1	1.2	1.4
$z_{s,3}$	0.78	0.94	1	1.2	1.4	1.6
σ^2_κ (10^{-6})	1.2	0.25	0.95	0.21	1.2	1.1

Table 3.5: Variance of the nulled convergence field fitted from the simulated data for various sets of source redshifts, 3 for each map, as labelled. The opening angle is $\theta = 10$ arcmin. These values are notably used in figure 3.23.

case in a scenario where nulling is applied since the variance depends both on the choice of source planes and the normalisation constant (which is totally arbitrary and could therefore be arbitrarily large!) which renders even less meaningful a direct interpretation of the width of the PDF in terms of level of non-Gaussianity. To do so, one would need to go back to projected densities (that is to say divide by $\kappa^{\rm null}_{\rm min}$).

Let me finally highlight that the LDT prediction presented here performs significantly better than a Gaussian as it allows us to model the non-zero higher order cumulants and therefore captures the tails of the PDF. Because our observable is non-Gaussian, the peak of the distribution (i.e the most likely value) and the mean differ significantly so that a Gaussian is always a bad fit to the PDF, not only in the tails but also around the peak (0 for a Gaussian while negative in the non-Gaussian case as the PDF is skewed towards underdensities). In addition, note that the LDT approach significantly outperforms a standard log-normal approximation for $\kappa + \kappa_{\rm min}$ and our formalism can be applied to jointly model the statistics of lensing convergence considered here and tracer densities (Friedrich et al., 2018), a situation where the log-normal approximation has been shown to fail.

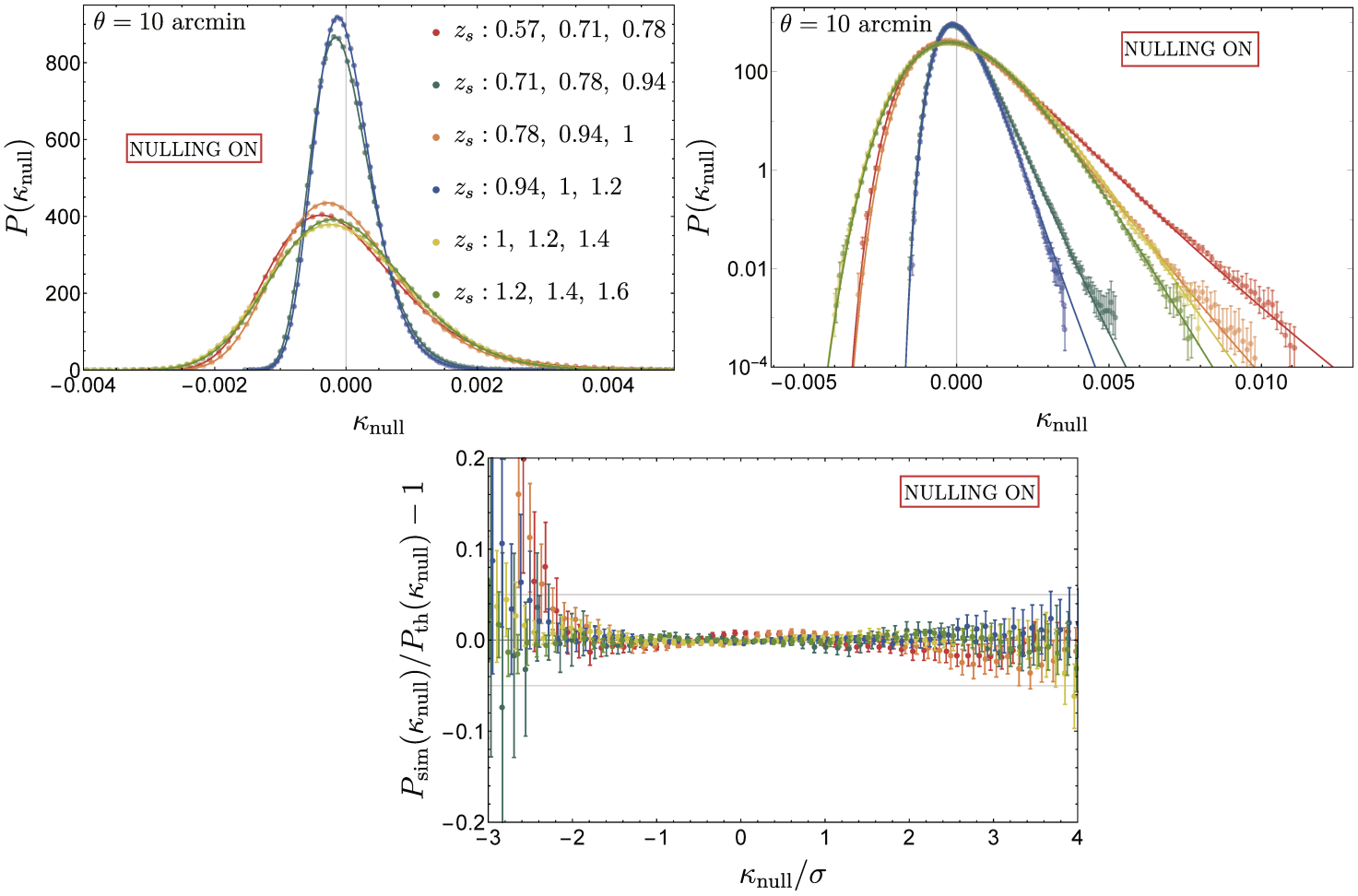


Figure 3.23: One-point PDF of the nulled weak-lensing convergence maps for different sets of redshift planes ordered as labelled from high variance (red) to lower variance (blue). The associated sets of planes leading to the nulled lensing kernel are roughly displayed in figure 1.12 (not the exact same source redshifts). The opening angle is fixed here to $\theta = 10$ arcmin. Solid lines display the LDT predictions while the measurements on the simulated sky are shown with error bars. Top left panel: PDF in linear scale. Top right panel: same as top panel in log scale to better display the tails. Bottom panel: residuals of the simulated data compared to the prediction. These plots have to be contrasted with figure 3.4 obtained before nulling. This figure was published in **Barthelemy** et al. (2020a).

Coming up with a realistic model :

Expectation ***Reality***

4

An attempt at a more realistic strategy: Aperture mass statistics and joint tomographic modelling

Contents

4.1. Introduction

Until now the results presented in this thesis have focused on a non-perturbative non-linear framework for the clustering of the matter density field by means of large deviation theory and used to compute the one-point statistics of any projection of matter, thus including the convergence field at first order in the potential. Additionally, I have also detailed the calculations of some of the lensing corrections on the skewness, such as but not limited to post-Born corrections, and presented a method to consistently include these corrections in the formalism. The cosmological information content of the convergence PDF was evaluated, and since projection effects tend to mix small and large physical scales I have used the so-called BNT transform, which allows us to narrow the kernel of lensing observables. However, although interesting in and of themselves and necessary for the following, the previous results suffer from two main issues, i) the convergence field is not a direct product of cosmic shear experiments although it could be obtained through magnification and ii) we know that experimental noise such as shape noise will explode for nulled quantities as a result of them being linear combinations of regular source-bins (we sum the noise in each bin and we get a less intense lensing signal resulting from fewer lenses). Indeed we both add the shape noise coming from each bin and the lensing signal itself becomes weaker since it results from the contribution of fewer objects. As such and in its current state, though being a neat theoretical tool, we will see that it is not usable for our PDFs if we stick to our current 1-nulled-bin analysis. This final chapter thus aims at answering, or at least begin to, those two main issues. The first one is addressed by moving from the convergence to the more observable aperture mass $M_{\rm ap}$ and the second one by presenting the whole formalism for joint PDFs of multiple nulled bins.

4.2. From the convergence to the aperture mass PDF

4.2.1 Definition and projection formula

As defined chapter 1, the aperture mass is seen as a compensated filtering of the convergence field. In principle, the large deviation formalism could be used with any filter function as was shown in Bernardeau & Reimberg (2016), Reimberg & Bernardeau (2018) and in section 4.4. However, I here adopt a simple prescription

$$M_{\rm ap}(\boldsymbol{\vartheta}) = \kappa_{<\theta_2}(\boldsymbol{\vartheta}) - \kappa_{<\theta_1}(\boldsymbol{\vartheta}), \tag{4.1}$$

where $\kappa_{<\theta_{1,2}}$ denotes the convergence field filtered by a top-hat window (see equation (1.67)) of angular radius θ_1 and $\theta_2 = 2\,\theta_1$. This choice is both motivated by the relative simplicity of obtaining statistics of concentric disks/spheres within the large deviation formalism – other choices of compensated filters are often used in the literature (see for starter Schneider et al. (1998)) – and also because top-hat filtering allows for a more rigorous assessment of the scales correctly described by the theory presented in this paper. We show the shape of our filter in figure 4.1 compared to the one used in Schneider et al. (1998).

Under the small-angle/Limber approximation, I have already shown that the convergence field filtered in an angular top-hat window function can be seen as a juxtaposition of statistically independent 2D slices of the underlying density field. Filtering the field using

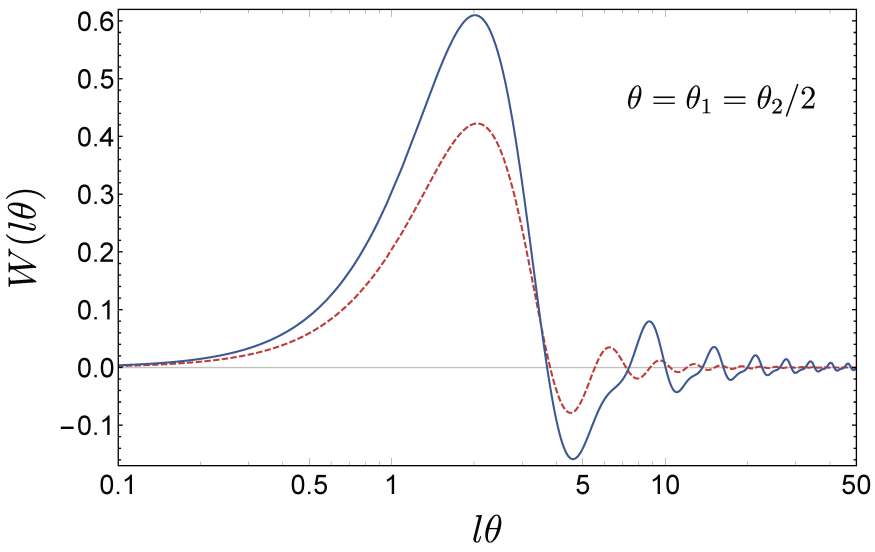


Figure 4.1: Comparison of the filter given in equation (4.1) (blue solid line multiplied by -1) and the one used in Schneider et al. (1998) (red dashed line). This figure was published in **Barthelemy** et al. (2021b).

equation (4.1) does not change the demonstration and thus the cumulants of the aperture mass lensing field are given by

$$\langle M_{\rm ap}^p \rangle_c = \int_0^{\chi_s} \mathrm{d}\chi \omega^p(\chi, \chi_s) \langle (\delta_{<\mathcal{D}(\chi)\theta_2} - \delta_{<\mathcal{D}(\chi)\theta_1})^p \rangle_c, \tag{4.2}$$

where $\delta_{<\mathcal{D}(\chi)\theta_2} - \delta_{<\mathcal{D}(\chi)\theta_1}$ is a random variable defining the density slope between two concentric disks of radii $\mathcal{D}(\chi)\theta_2$ and $\mathcal{D}(\chi)\theta_1$ at comoving radial distance χ. Equation (4.2) thus reduces the complexity of the problem down to computing the one-point statistics of the density slope in each two-dimensional slice (equivalently the slope between infinitely long cylinders at the same redshift) along the line-of-sight. The one-point statistics of the 2D density slope will be obtained via large-deviation theory and I will then build the non-linear cumulant generating function of the aperture mass and its PDF. A schematic representation of our procedure can be found in figure 4.2.

4.2.2 Non-linear $M_{\rm ap}$ CGF

When extending the results of large deviation theory to finite, non-zero values of the driving parameter, one usually just takes the driving parameter, here the variance of the density field, as a free parameter, and because it has in our case a physical meaning we set it to its "real", non-linear, value. This construction is indeed appealing in our cosmological context since it ensures that the reduced cumulants of the density field are exactly those obtained through standard Eulerian tree-order perturbation theory and it is this construction that we roughly kept in the previous chapter 3 for projections of the density field.

However, at the level of the rate function, this re-scaling by the driving parameter could be read as a simple modelling of the non-linear variance with a pivot scale as

$$\sigma_{nl}(R\rho^{1/2}) = \frac{\sigma_{nl}(R)}{\sigma_l(R)} \sigma_l(R\rho^{1/2}), \tag{4.3}$$

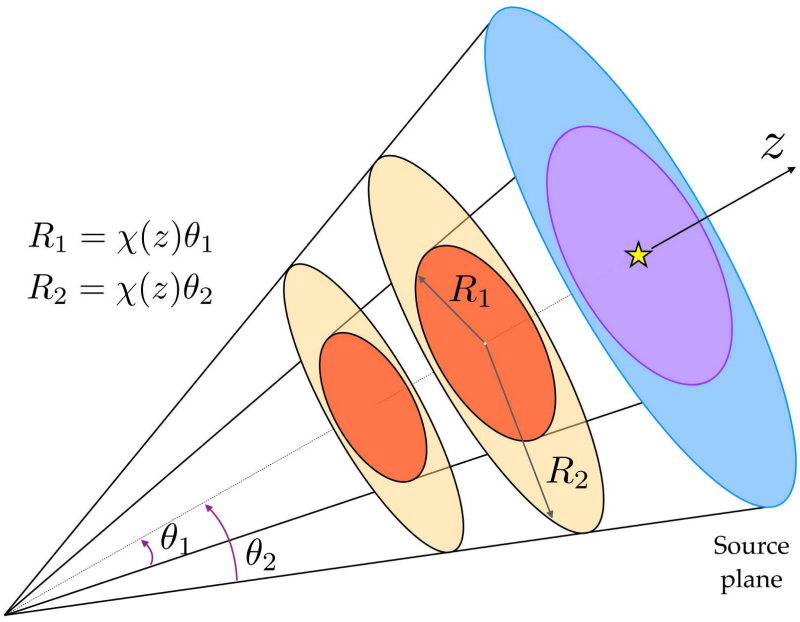


Figure 4.2: Schematic view of the procedure to predict the aperture mass one-point statistics. Here projected quantities are seen as a superposition of thin and statistically independent slices of the underlying 3D density field along the line of sight (equation (4.2)). Since the dynamics of disks inside a slice is on average well-described by cylindrical collapse, we use it to study the joint statistics of the 2D density field at two different scales (yellow and orange) at every redshift up to the source plane. The scales studied at each redshift allow me to take into account the geometry of the light-cone. This figure was published in **Barthelemy** et al. (2021b).

where $\sigma^2_{nl}(R)$ is the driving parameter taken in its non-zero, finite value. As a matter of fact there is not, to my knowledge, any mathematical prescription for extending the large deviation results to the non-zero limit. Moreover, for the case of the joint statistics of the density field at different scales, the choice of a unique driving parameter prevents us from imposing all the correct quadratic contributions in the CGF, which is particularly problematic for the one-point PDF of the aperture mass. Note indeed that as previously seen for the convergence, sums of random variables satisfying a large deviation principle do not necessarily satisfy the same principle. As a consequence, I here choose to model the non-linear covariance, not by a re-scaling by the driving parameter, which thus do not matter in the final expression of the CGF, but by a full non-linear prescription that comes in our case from Halofit. This ensures that all quadratic contributions in the CGF are correctly modelled and modifies the standard tree-order PT results, keeping their functional form but using a non-linear power spectrum where the linear one usually appears. To assess the further validity of this expression in terms of perturbation theory, one would then need to compare these results to high-order loop calculations and check the meaning of these corrections. This goes beyond this thesis' results and is left for future work.

As another sanity check to ensure that any possible discrepancies with the numerical simulation are not influenced too much by this choice, we also re-scale the projected CGF by the measured variance $\sigma^2_{M_{\rm ap},\rm sim}$ instead of the one computed with Halofit $\sigma^2_{M_{\rm ap},\rm hfit}$,

$$\phi_{M_{\rm ap}}(\lambda) = \frac{\sigma^2_{M_{\rm ap},\rm hfit}}{\sigma^2_{M_{\rm ap},\rm sim}}\phi_{M_{\rm ap}}\left(\lambda\frac{\sigma^2_{M_{\rm ap},\rm sim}}{\sigma^2_{M_{\rm ap},\rm hfit}}\right), \tag{4.4}$$

so that our resulting $M_{\rm ap}$ CGF contains the exact variance and all higher-order cumulants given by the non-linear collapse dynamics. Note that, contrary to what we could for the convergence field, we cannot take the linear covariance along the line of sight and re-scale everything by the $M_{\rm ap}$ variance. Indeed, it would still impose the need for a driving parameter for the density slope in each slice, though taken to its linear value, which thus changes all quadratic contributions in the CGF of each slice. Then the re-scaling by the non-linear $\sigma^2_{M_{\rm ap}}$ would correct the variance but lead to very spurious high-order cumulants. I give an example of different treatments of the covariance at the denominator of the rate function along the line of sight in the next subsection.

As a consequence, we cannot as for the convergence keep only one "free" parameter. Eventually, this comes with the modelling of a true non-Gaussian observable (up to a reduced shear correction but as opposed to the convergence which needs reconstruction of mass maps), which is sensitive to multiple scales at once, a property usually very useful to help break degeneracy between cosmological parameters as we saw in the previous chapter 3.

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

$$\sigma^2(R_1,R_2;z) = \int \frac{{\rm d}^2\boldsymbol{k}_\perp}{(2\pi)^2}P(k_\perp;z)W_{TH}(R_1k_\perp)W_{TH}(R_2k_\perp), \tag{4.5}$$

where $W_{TH}(l) = 2J_1(l)/l$ and J_1 is the first Bessel function of the first kind; ii) this enables us to compute the rate function for any values of the densities inside each disk in a given

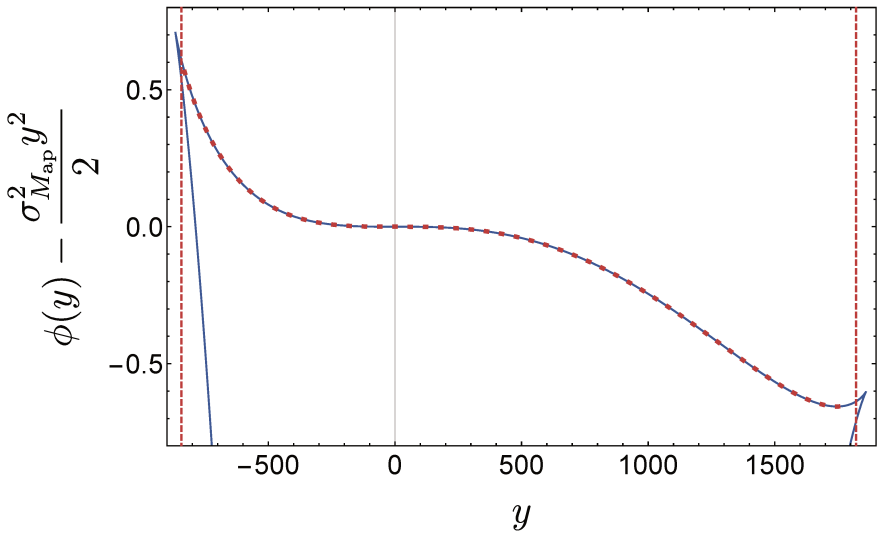


Figure 4.3: Cumulant generating function of the aperture mass at redshift $z_s = 1$, $\theta_1 = 15$ and $\theta_2 = 30$ arcmin. The dashed red line is the CGF as computed with large deviation theory and the blue line is the one computed with a fitted effective collapse of degree 7. The dashed red vertical lines indicate the position of the critical points of the CGF as computed with LDT. The successive derivatives in zero (cumulants) are perfectly reproduced, as well as the general shape and location of critical points. The $\sigma^2_{M_{\rm ap}} y^2/2$ term was substracted here to remove the quadratic contribution to the CGF and better display the part, corresponding to high-order cumulants, modelled by our formalism. We subtract the exact same quantity to the 2 curves so that the agreement displayed here is not affected by that choice. This figure was published in **Barthelemy et al. (2021b)**.

slice along the line of sight; iii) numerically inverting the stationary condition, we can now compute for any real λ the CGF of the 2D density slope within each slice; iv) using the projection formula (4.2) and equation (4.4) we finally compute the non-linear CGF of the aperture mass. To perform the analytical continuation of the CGF in the complex plane, which is necessary to compute the PDF, I use the effective collapse method already described in subsection 3.5.3.

Overall figure 4.3 illustrates how well this approach reproduces the $M_{\rm ap}$ cumulant generating functions computed with large deviation theory. In particular for this procedure, being from a mathematical point of view strictly identical to the usual large deviation approach applied to matter density field one-point statistics, the critical behaviour that the (S)CGF exhibits along the real axis and that is the result of a change of convexity of the rate function or equivalently multiple solutions to the stationary condition (which are apparent on the blue line of figure. 4.3), will also be present in the case of reconstructed generating functions via an effective mapping. Finally, recall that though the procedure we described in this section is the one that we implement to construct the aperture mass PDF, I should technically use this construction at the level of every slice along the line of sight as discussed in subsection 3.5.3.

4.2.3 $M_{\rm ap}$ and BNT $M_{\rm ap}$ PDFs

As an illustration, figure 4.4 displays the resulting prediction for an aperture mass PDF for a single source redshift at $z_s = 1$. The opening angle is chosen to be $\theta_1 = 15$ arcmin

and the non-linear covariance of at the denominator of the rate function in each redshift slice is treated in three different ways, namely i) the linear prediction with a re-scaling of the obtained SCGF by a non-linear driving parameter – the variance of the field at the smallest scales – inside each redshift slice along the line of sight, ii) using the full Halofit power spectrum as input [our baseline approach] and iii) using the Euclid emulator of the non-linear power spectrum (Euclid Collaboration et al., 2019) for comparison of different non-linear prescriptions. The blue solid line shows that indeed re-scaling the SCGF by the driving parameter as in the 1-cell case does not lead to the correct non-linear variance of the aperture mass (we assume in this subsection that it is equal to the Halofit/Euclid emulator prediction), which would lead to a major source of disagreement between this model and the measured/simulated PDF. This further illustrates the discussion of the previous subsection where the importance of using the full non-linear power spectrum was underlined. However, the $M_{\rm ap}$ PDF does not appear to be very sensitive to the precision in the modelling of the non-linear power spectrum as the very good agreement between the Halofit and Euclid Emulator prescriptions implies. Indeed, and as illustrated in figure 8 of Euclid Collaboration et al. (2019), the scale-dependence of the two power spectra is not exactly the same though this does not seem to affect significantly the values of the high-order cumulants in the PDF of the aperture mass.

As expected from our formalism, we observe two exponential cut-offs on each side of the PDF, the positive tail being driven by large values of the convergence field filtered at the scale θ_2 and the negative tails by large values of the convergence field filtered at the scale θ_1. With the convergence field becoming more and more skewed with decreasing smoothing scale, we expect that the negative tail of the aperture mass is the most prominent one, which also implies a shift in the most likely value towards positive values. However note that those considerations are only rough approximations that help us to understand the general shape of the aperture mass PDF. In particular, the tails of the $M_{\rm ap}$ PDF are not at all identical – different power laws – to the respective tails of $P(\kappa_{<\theta_1})$ and $P(\kappa_{<\theta_2})$, which are the tails of the convolution of the 2 convergence PDFs if the two scales were independent. Overall the displayed PDF is far from Gaussian, thus highlighting how non-negligible non-Gaussian features of the aperture mass distribution are in this regime.

The BNT transform is as straightforward to implement for the aperture mass as it is for the convergence since we once again only need to replace the original kernel with its nulled counterpart. I show in figure 4.5 how this construction allows for a very effective description of the $M_{\rm ap}$ one-point statistics by comparing our formalism – case ii) of figure 4.4 – to measurements made in the numerical simulation described in the following section. One can appreciate that the exponential cut-off in the tails of the PDF, a prediction of our formalism, is well-observed once one reduces the lensing kernel down to scales accessible to first-principles theoretical modelling (*i.e.* perturbation theory). Apart from the general shape, one can also see that the theoretical PDF agrees really well with the measured one, way below the percent in the bulk and within at least 5% in the $\pm$ 4σ region around the peak.

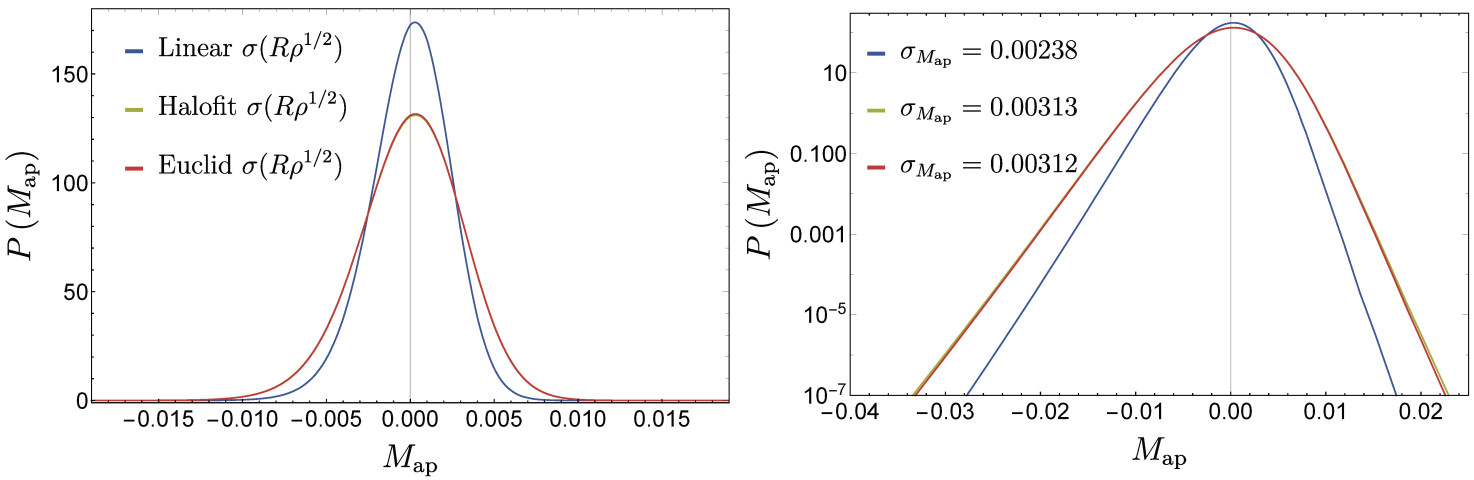


Figure 4.4: Aperture mass PDF at $z_s = 1$, $\theta_1 = 15$ and $\theta_2 = 30$ arcmin. PDFs are obtained fitting an effective collapse of degree 7 and we compare different prescriptions for the non-linear covariance of disks in redshift slices along the line of sight. The blue line is the traditional 1-cell approach where the CGF computed with the linear covariance is re-scaled by the non-linear driving parameter. The green and red solid lines are the PDF obtained from the full non-linear CGF with different prescriptions for the non-linear covariances. The green and red curves are almost indistinguishable which illustrates that the $M_{\rm ap}$ PDF is not sensitive to the detailed modelling of the non-linear matter power spectrum. This figure was published in **Barthelemy** et al. (2021b).

4.3. Simulated $M_{\rm ap}$ cumulants: example of the need for a theoretical model

In principle, I would expect the large deviations and nulling formalism presented in this thesis to perform extremely well as the BNT transform localises the lensing kernel to a finite range of redshifts and therefore physical scales thus making the $M_{\rm ap}$ one-point statistics closer to the density slope in long cylinders for which large deviation formalism was already proven very efficient for example in Uhlemann et al. (2018c). Moreover, this strategy applied to convergence maps was also found to be very effective. Also note that a very satisfying agreement between the theory and this simulation suite was already found in figure 4.5 and that I would merely like to extend this result down to smaller scales and different ratios between the filtering scales.

The difficulty that has been encountered when trying to assess the validity regime of our theoretical approach lies in the difficulty in precisely measuring and assigning error bars to measured $M_{\rm ap}$ quantities. In the context of the PDF, this can be further exemplified by the measurement of the skewness, a single number but that determines the first and foremost (thanks to the cumulant hierarchy) non-trivial contribution to non-Gaussianity in the PDF. To this end, I study the BNT $M_{\rm ap}$ skewness with opening angles of $\theta_2 = 15$ and $\theta_1 = 10$ arcmin and source planes respectively located at redshift $z_s = 1.2 - 1.4 - 1.6$. For 3 different realisations of the full sky maps – except for the greatest resolution which only offers a single realisation – I apply our filtering scheme, measure the BNT $M_{\rm ap}$ skewness in the map and measure error bars as standard error on the mean computed among 8 subsamples of the full sky. I find that: i) the 2 different filtering schemes (in real and harmonic space) give similar

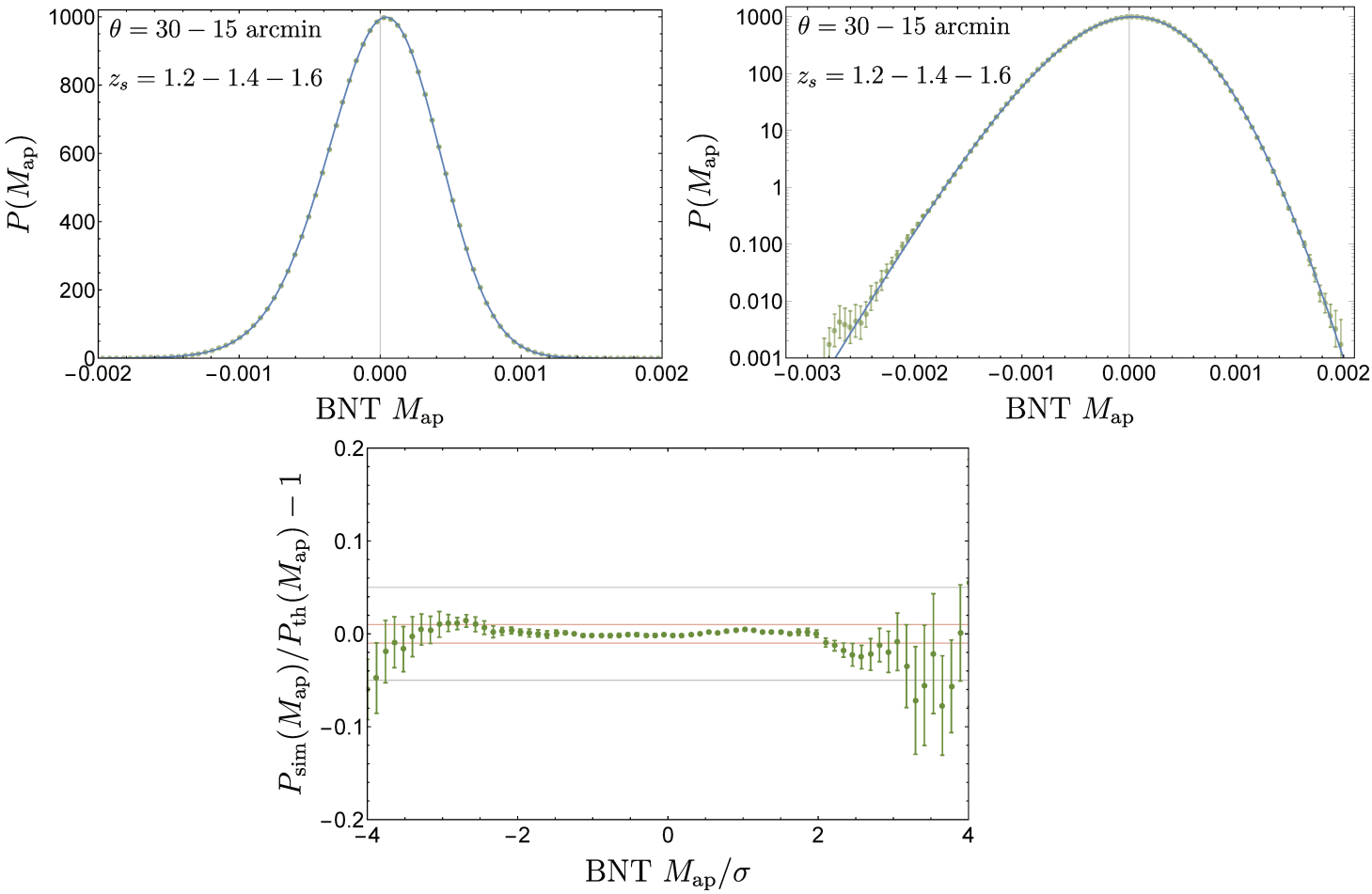


Figure 4.5: Theoretical BNT $M_{\rm ap}$ PDF compared to one measured in the Takahashi simulation. The expected exponential cut-offs in the tails are well observed when one restricts the lensing kernel to physical scales accessible to perturbation theory. The hereby formalism also agrees very well with the measured PDF (from one realisation of the map at the lowest resolution). Note that the error bars represent the standard error-on-the-mean computed from 8 sub-samples of the full-sky. The red horizontal lines indicate the region of ± 1% residual and the grey ones ± 5%. This figure was published in **Barthelemy** et al. (2021b).

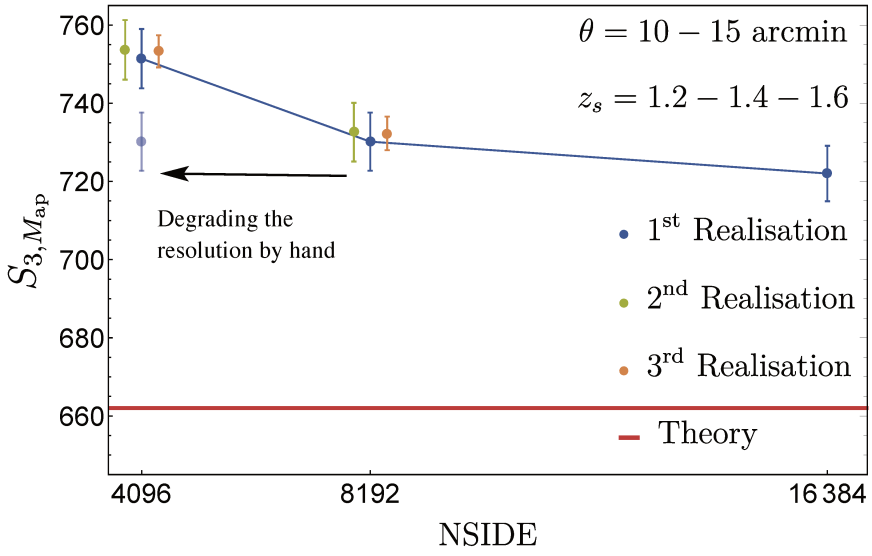


Figure 4.6: Different BNT $M_{\rm ap}$ skewness values as measured in the simulation as a function of the map resolution NSIDE for three different realisation with respectively blue, green and orange error bars. For comparison, the corresponding prediction from our formalism is displayed using a red solid line. This figure was published in **Barthelemy** et al. (2021b).

results with most of the time much more than percent precision. As shown on figure 4.6, ii) different realisations at the same resolution agree very well being statistically compatible with one another as probed by the error bars and with mean values very close to each other; iii) there is a visible shift of the value with increasing resolution to the point that the lowest and highest resolution – though coming from the same realisation – do not seem compatible, which also does not seem to come from resolution itself since degrading the map by hand to a lower resolution, filtering it and then re-measuring the skewness leads to very similar results; iv) none of the measured values agree with the theoretical prediction ($\lesssim 10\%$ relative difference). This is reminiscent of for example section 4.2 of Uhlemann et al. (2019) where the impact of resolution of the N-body simulation observed for the reduced skewness of the 3D matter density field was similar though the two cases are not exactly comparable. Also note that to some extent, those observations are still valid in the case presented previously in figure 4.5 but smaller. There, the PDF was measured from one single realisation of the map at the lowest resolution and proved to perform very well.

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

A more careful examination of the BNT $M_{\rm ap}$ skewness value can also be performed decomposing it as cross-cumulants of the BNT convergence field filtered at two different scales. It is written as

$$S_{3,M_{\rm ap}} = \frac{\langle M_{\rm ap}^3 \rangle_c}{\sigma_{M_{\rm ap}}^4} = \frac{\langle \kappa_2^3 \rangle_c - \langle \kappa_1^3 \rangle_c + 3\langle \kappa_1^2 \kappa_2 \rangle_c - 3\langle \kappa_1 \kappa_2^2 \rangle_c}{\sigma_{M_{\rm ap}}^4}. \tag{4.6}$$

I then measured each cross-cumulant in the first realisation at the lowest and medium resolutions and compared them to their respective theoretical predictions as shown in figure 4.7.

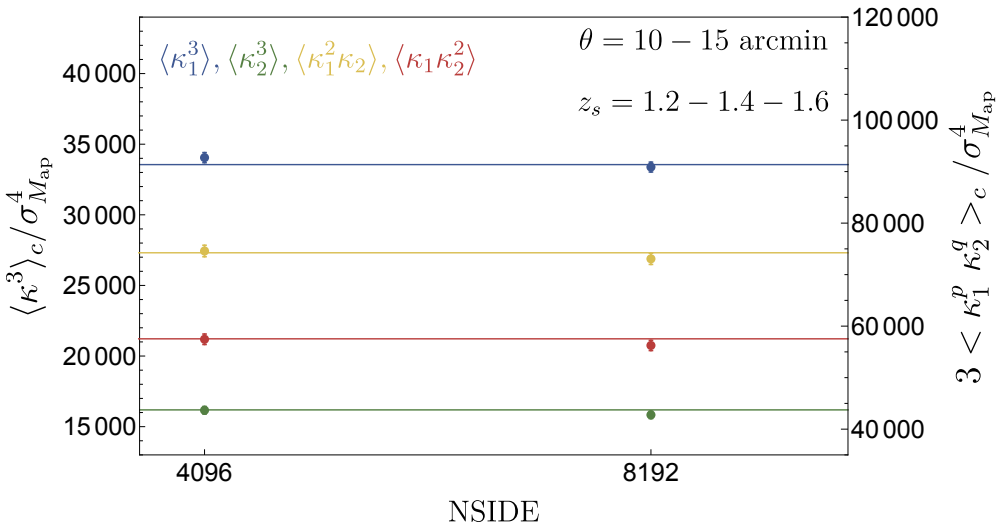


Figure 4.7: Reduced cross-cumulants of the nulled convergence field appearing in equation (4.6). Points with error bars are measured in the simulation's first realisation and solid lines represent the corresponding theoretical predictions. Even though the values tend to be quite sensitive to the resolution, a sub-percent agreement with the theoretical predictions is found. This figure was published in **Barthelemy** et al. (2021b).

The agreement for those cross-cumulants is found to be excellent (to the percent precision and very close to the 1-σ error bar), allowing us to extend the results obtained on the one-point convergence to a multi-scale analysis. Let us notice that in equation (4.6) numbers of similar orders of magnitude are subtracted from each other, which in turn tends to decrease the precision of the prediction of the BNT $M_{\rm ap}$ skewness since the absolute difference tends to stay the same while the subtraction of two numbers of similar magnitude results in a smaller number, which finally increases the relative difference between the theory and the numerical simulation. We are thus looking at some very subtle effects in the convergence field itself that tend to have a significant impact on the quantities we are interested in. This could be the sign of higher order perturbative corrections (loop corrections) but given that observables similar to the BNT $M_{\rm ap}$ such as densities in cylinders were already studied and found to be accurate at tree-order (Uhlemann et al., 2018c), a more likely explanation is to be found on the simulation side. Indeed subtle numerical artefacts could show up in the BNT $M_{\rm ap}$ such as resolution or discretisation effects, convergence of the N-body algorithm etc. In fact, since the relative incoherences observed in the skewness when increasing the resolution of the maps are not seen at all in the convergence field itself, it could indeed be possible that small errors in the simulation are amplified when looking at $M_{\rm ap}$ statistics.

Finally note that I also tried to perform – sometimes successfully – comparisons of our theoretical formalism to other numerical suites. However, other issues such as the evaluation of the degree of independence between realisations of randomised lines of sight in replications of a small N-body box, and other simulation-specific issues prevented me from doing any more refined comparison to available ray-tracing simulations. There is therefore a dire need for future code comparisons, validations and improvements in the line of Hilbert et al. (2020) that specifically target higher-order statistics. Such future developments could then allow us to test in more detail the validity regime of our theoretical model for the BNT $M_{\rm ap}$ statistics. In any case, exploitation of the $M_{\rm ap}$ PDF as an observable should not, in light

of these findings, be done assuming that numerical simulations are sufficiently accurate, and physical formalisms just like this one should be considered to at least test the different numerical schemes in cosmological analyses of non-Gaussian statistics.

Resolution effects: Before concluding, let me mention two additional tests that were done on the theory side to try to mend the discrepancy between the theory and the simulation. I separate these since they are not tied at all to the comparison to simulations in general. First, I tried to introduce an l-cut due to the resolution in the computation of the skewness itself which is done by introducing the corresponding k-cut at each slice along the line of sight thus modifying the variance and its derivatives. Second, I considered the effect of having a non-exact top-hat filter in section 4.4 choosing the L parameter – a measure of "how smoothed" the top-hat with an exact top-hat given for $L \to \infty$ is, where typically $L = 20$ for the lowest resolution – to smooth the filter over 2 pixels. Those two prescriptions were shown to have no significant impact on the skewness and therefore are unlikely explanations for the observed differences.

4.4. Going beyond top-hat smoothing

One issue of top-hat filtering is its "sharpness" (non differentiability at the radius) which could make its precise implementation to real, pixelated data challenging. Hence the purpose of this section is to test the sensitivity of the predicted value of the aperture mass skewness to a slight modification of its filter by using a smoothed C^∞ version of the top-hat defined by

$$F(x, L, R) = \frac{1}{2}\left(1 + \mathrm{erf}\left[L\left(1 - \frac{x}{R}\right)\right]\right)/V(L, R), \tag{4.7}$$

where R is the radius of the top-hat, L is a parameter that influences the sharpness of the filter – we recover a top-hat in the limit where L tends to infinity – and $V(L, R)$ is the normalisation of the filter given by

$$V(L, R) = \pi R^2 \frac{\left(2e^{-L^2}/\sqrt{\pi} + (1 + 2L^2)(1 + \mathrm{erf}(L))\right)}{4L^2}. \tag{4.8}$$

I show in figure 4.8 the shape of the filter depending on the value of its sharpness parameter L. Finally note that equation (4.7) is built so that the departure from an actual top-hat depends solely on L and not on the radius so that the width needed for the filter to go from its maximum value to 0 is i) centred around the radius and ii) a fixed percentage of its value for a given L. Roughly, $L = 20$ corresponds to $\pm 10\%$ around R, $L = 40$ is $\pm 5\%$ and so on linearly.

The principle behind the implementation of any filter using large deviation theory lies in its expression as a linear combination of top-hat kernels taken in the continuous limit (Bernardeau & Reimberg, 2016; Reimberg & Bernardeau, 2018). First let me notice that since

$$\kappa_{<\theta} \equiv \int_0^\theta \frac{\mathrm{d}^2\boldsymbol{\vartheta}}{\pi\theta^2}\kappa(\boldsymbol{\vartheta}) \tag{4.9}$$

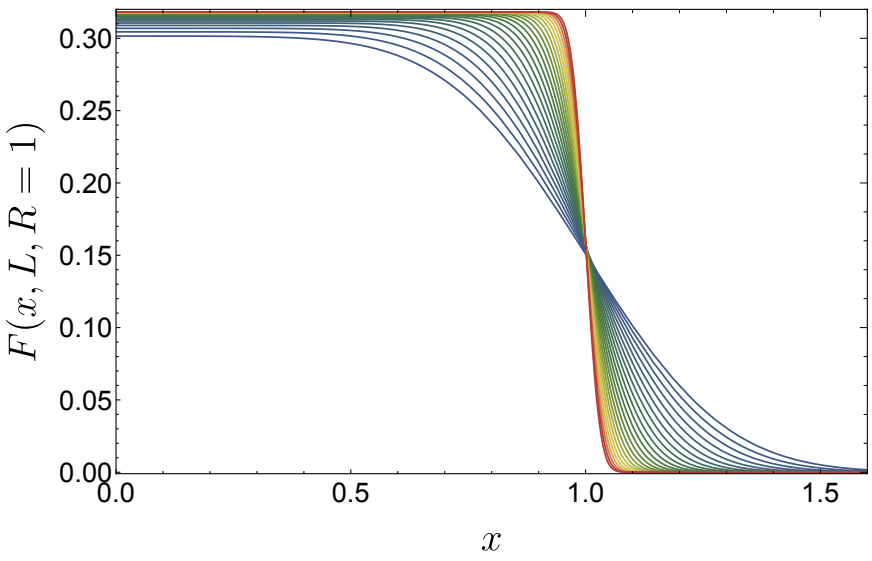


Figure 4.8: Smoothed top-hat window function for values of L from 3 (blue) to 30 (red) and $R = 1$. This figure was published in **Barthelemy** et al. (2021b).

then I obtain, by differentiating,

$$\kappa(\vartheta) = \kappa_{<\vartheta} + \frac{\vartheta}{2}\kappa'_{<\vartheta}. \tag{4.10}$$

Thus for any filter defining some quantity $\hat{\kappa}$ we get by integration by parts

$$\hat{\kappa}_{=} \int \mathrm{d}^2\boldsymbol{\vartheta}\, U(\vartheta)\kappa(\boldsymbol{\vartheta}) = \int \mathrm{d}\vartheta\, \hat{U}(\vartheta)\kappa_{<\vartheta} \tag{4.11}$$

with

$$\hat{U}(\vartheta) = -\pi\vartheta^2 U'(\vartheta). \tag{4.12}$$

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

$$\hat{U}_{M_{\rm ap}}(\vartheta) = \frac{4\vartheta^2 e^{L^2} L^3}{\sqrt{\pi} e^{L^2}\left(2L^2+1\right)\left(\mathrm{erf}(L)+1\right) + 2L} \times \left(\frac{\exp(-L^2(\theta_2-\vartheta)^2/\theta_2^2)}{\theta_2^3} - \frac{\exp(-L^2(\theta_1-\vartheta)^2/\theta_1^2)}{\theta_1^3}\right) \tag{4.13}$$

is shown in figure 4.9. Note that $\hat{U}_{M_{\rm ap}}$ would be a difference of two Dirac delta functions if we were using regular top-hats as is done in the previous sections. For simplicity I also use the same value of L for the two different scales which leads to broader $\hat{U}_{M_{\rm ap}}$ as θ_i increases, though this might not be an important issue for radii relatively close one to another.

I now need to express the SCGF of the density in each slice along the line of sight. This boils down to writing the continuous limit of Varadhan's theorem and the definition of the rate function re-written as

$$\varphi(\lambda) = \sup_{\{\rho_i\}} \left[\lambda \sum_i^N \hat{U}_i\rho_i - \frac{\sigma^2\left(R_N\right)}{2} \sum_{k,j} \Xi_{kj}(\{\tau_i\})\tau_k\tau_j\right]. \tag{4.14}$$

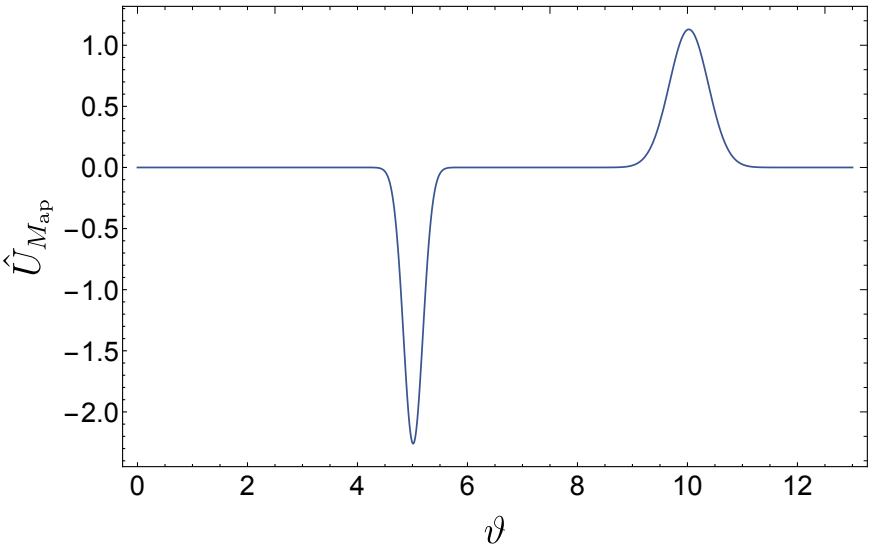


Figure 4.9: $\hat{U}_{M_{\rm ap}}$ for $\theta_1 = 5$, $\theta_2 = 10$ and $L = 20$. The departure from the exact top-hat for this L is roughly $\pm 10\%$ around the radius. This figure was published in **Barthelemy** et al. (2021b).

This is done by writing

$$\sum_{i=1}^{N} \hat{U}_i\left(R_i\right) \rho_{R_i} \rightarrow \int \mathrm{d}R \hat{U}(R) \zeta\left(\tau_r\right) = \int \mathrm{d}r \frac{\mathrm{d}R}{\mathrm{d}r} \hat{U}\left[R(r)\right] \zeta\left(\tau_r\right) \tag{4.15}$$

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

$$r = R\rho^{1/2}. \tag{4.16}$$

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

$$\int \mathrm{d}r' \sigma^2\left(r, r'\right) \xi\left(r', r''\right) = \delta_{\rm D}\left(r - r''\right). \tag{4.17}$$

Therefore the continuous limit to the SCGF can be written as

$$\varphi(\lambda) = \sup_{\tau} \left[\lambda \int \mathrm{d}r \frac{\mathrm{d}R}{\mathrm{d}r} \hat{U}\left[R(r)\right] \zeta\left(\tau_r\right) - \frac{\sigma^2_{\rm slice}}{2} \int \mathrm{d}r \mathrm{d}r' \tau_r \tau_{r'} \xi(r, r') \right], \tag{4.18}$$

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

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

$$S_3^{\rm slice} = 3\nu_2 \frac{\int \mathrm{d}x \hat{U}(x) \Sigma^2(x)}{\left[\int \mathrm{d}x \hat{U}(x) \Sigma(x)\right]^2} + 3 \frac{\int \mathrm{d}x x \hat{U}(x) \Sigma(x) \Sigma'(x)}{\left[\int \mathrm{d}x \hat{U}(x) \Sigma(x)\right]^2}, \tag{4.19}$$

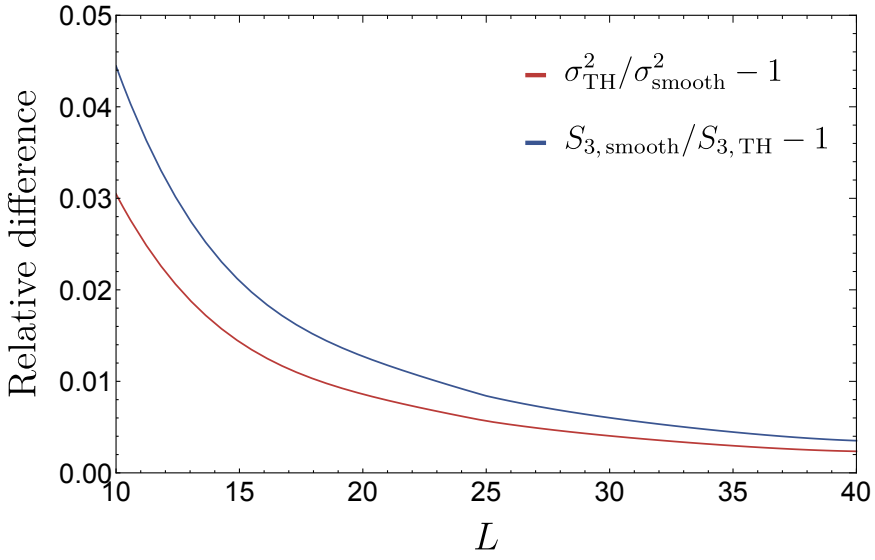


Figure 4.10: Red line: Relative difference between the $M_{\rm ap}$ variance coming from exact top-hat windows and using its smoothed counterpart in equation (4.7). Blue line: Same thing but for the $M_{\rm ap}$ reduced skewness. $\sigma^2_{\rm TH} = 2.17\ 10^{-7}$ and $S_{3,\rm TH} = -596.08$. This figure was published in **Barthelemy** et al. (2021b).

where the coefficient in front of the second term is actually six over the dimension of the collapse, ν_2 is the usual spherical collapse coefficient and we have

$$\Sigma(x) = \int \mathrm{d}y \sigma^2(x,y)\hat{U}(y). \tag{4.20}$$

Now combining the result in equation (4.19) and the projection formula one can get to the aperture mass skewness taking into account the geometry of the light-cone as well as the smoothed top-hat filter (4.7). For a nulled aperture mass field with source redshifts located at $z_s = 1.2 - 1.4 - 1.65$ and smoothing angles $\theta_1 = 10$ and $\theta_2 = 20$ arcmin, I compute both the variance and the reduced skewness using an exact top-hat window and its smoothed version for different values of L. I also use a Halofit power-spectrum. The relative difference between the two is displayed in figure 4.10, where one can see that the values of both the variance and skewness depend very weakly on the value of L. This is overall good news since approximate filtering schemes could thus be considered with still relatively good validity of our theoretical modelling.

4.5. Shape noise on individual nulled lens-bins

Being obtained from the convergence of the tangential shear, the aperture mass is still subject to the variation of the intrinsic ellipticity of galaxies known as shape noise. Still modelling it as an added Gaussian of zero mean and variance σ_{SN} on the convergence field, the resulting Gaussian noise of the aperture mass assuming the same number of observed galaxies in each source-bin/source-plane is straightforwardly expressed as

$$\sigma^2_{SN,M_{\rm ap}} = \frac{\alpha\sigma^2_\epsilon}{\pi n_{g_s}}\left(\frac{1}{\theta_1^2} - \frac{1}{\theta_2^2}\right), \tag{4.21}$$

where $\sigma_\epsilon^2 = 0.3$ is the Gaussian noise on the determination of the lensing-induced ellipticity of each galaxy, $n_{g_s} = 3$ arcmin^{-2} is the normalised mean number of observed galaxies, $\alpha = 1$ for regular $M_{\rm ap}$ and $\alpha = \sum_{i=j-2}^{j} \left(M^{ij}\right)^2$ for BNT $M_{\rm ap}$. The resulting noisy theoretical $M_{\rm ap}$ PDF is then obtained by convolving the previously computed PDFs with a zero-mean Gaussian of the appropriate variance

$$P_{SN}(M_{\rm ap}) = \frac{1}{\sigma_{SN,M_{\rm ap}}\sqrt{2\pi}} \int_{-\infty}^{+\infty} {\rm d}\hat{M}_{\rm ap} P(\hat{M}_{\rm ap}) \exp\left(-\frac{(M_{\rm ap}-\hat{M}_{\rm ap})^2}{2\sigma^2_{SN,M_{\rm ap}}}\right). \tag{4.22}$$

As an illustration of the effect of shape noise on the $M_{\rm ap}$ PDF I show in figure 4.11 the relative difference between the noisy PDFs $P_{SN}(M_{\rm ap})$ and a Gaussian of the same mean and total variance (noise + signal). This is done by adding Gaussian noise to each pixel of the simulated convergence maps of 10 realisations of the Takahashi simulation and still at redshifts mimicking a Euclid binning, combining them so as to get nulled maps, measuring the resulting $M_{\rm ap}$ PDFs and computing the standard deviation between the 10 realisations in each bin of the $M_{\rm ap}$ PDF as an estimate of the cosmic variance. We fix the opening angle to $\theta_2 = 2\theta_1 = 30$ arcmin and show both the residuals of the noisy $M_{\rm ap}$ PDF with respect to a Gaussian of the same variance for source redshifts $z_s = 0.57$ and 1.2 and nulled bins in between. Note that, as an additional tool, I also use the quick method based of the variance of the skewness estimator presented in the previous chapter 3 to estimate the detectability of non-Gaussian features in the BNT Aperture mass PDF in the presence of shape noise.

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

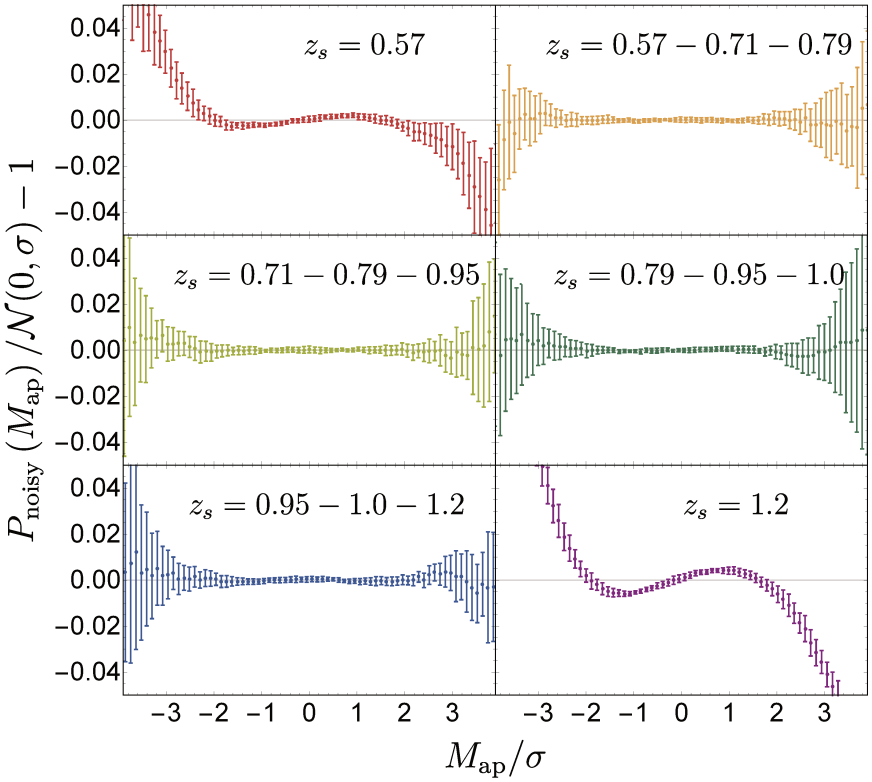


Figure 4.11: Residuals of the noisy $M_{\rm ap}$ PDFs with respect to a Gaussian of the same variance. The opening angle is $\theta_1 = 15$ and $\theta_2 = 30$ arcmin. The regular $M_{\rm ap}$ fields are taken at source redshifts $z_s = 0.57$ and 1.2, and the BNT $M_{\rm ap}$ fields have source redshifts located in between mimicking a Euclid binning. Error bars are taken as the standard deviation between 10 realisations of the simulation with added noise to estimate the cosmic variance. This figure was published in **Barthelemy** et al. (2021b).

4.6. Joint PDF of BNT lens-bins

4.6.1 Joint nulled CGFs

In order to predict the joint statistics in multiple redshift bins, the starting point is the expression of the cumulant generating functions of projected densities, which will, in our case, correspond to the convergence or the aperture mass. Once again, considering that each redshift slice along the line of sight is independent of the others I get that

$$\phi(\{\lambda_i\}) = \int \mathrm{d}r\, \phi_{\mathrm{cyl/slope}}\left(\sum_i \lambda_i\, w_i(r); r\theta_0; z(r)\right) \tag{4.23}$$

where θ_0 is the filtering angular size, $w_i(r)$ is the line-of-sight profile which leads to the observable κ_i map, $\phi_{\mathrm{cyl.}}(\lambda; R; z)$ is the cumulant generating function of the density in a cylinder of transverse size R, at redshift z, written as a function of λ, and $\phi_{\mathrm{slope}}(\lambda; R; z)$ is the cumulant generating function of the difference between the densities in two concentric cylinders of transverse sizes $R \equiv R_1$ and $R_2 > R_1$ at redshift z. The subscripts cyl and slope respectively represent the convergence and the aperture mass and I write $\phi_{\mathrm{cyl/slope}}$ when either or the other can be used. For simplicity, I will present the formalism for the convergence κ but the same equations hold for the aperture mass field albeit replacing cyl by slope.

The joint probability density function (PDF) of the maps κ_i corresponding to redshift bins indexed by i can then be constructed through an inverse Laplace transform of the cumulant generating function,

$$\mathcal{P}(\{\kappa_i\}) = \int \prod_{i=1}^{n_t} \frac{\mathrm{d}\lambda_i}{2\pi\,\mathrm{i}} \exp\left[-\sum_{i=1}^{n_t} \lambda_i\kappa_i + \phi(\{\lambda_i\})\right] \tag{4.24}$$

where n_t is the total number of source planes and where the integrals run along the imaginary axis and simply note that marginalizing over one variable κ_p is obtained by setting the corresponding variable λ_p to 0,

$$\begin{aligned}\int \mathrm{d}\kappa_p \mathcal{P}(\{\kappa_i\}) = \int \frac{\mathrm{d}\lambda_1}{2\pi\,\mathrm{i}} \cdots \frac{\mathrm{d}\lambda_{p-1}}{2\pi\,\mathrm{i}} \frac{\mathrm{d}\lambda_{p+1}}{2\pi\,\mathrm{i}} \cdots \frac{\mathrm{d}\lambda_{n_t}}{2\pi\,\mathrm{i}} \times \\ \exp\left[-\sum_{i=1}^{p-1} \lambda_i\kappa_i - \sum_{i=p+1}^{n_t} \lambda_i\kappa_i + \phi(\{\lambda_i\}|_{\lambda_p=0})\right].\end{aligned} \tag{4.25}$$

This property will be very useful in the following.

Let me now assume that all sources are located on discrete source planes and consider the BNT transform

$$\hat{w}_i(r) = \sum_j p_{ij}\, w_j(r) \tag{4.26}$$

in such a way that the functions $\hat{w}_i(r)$ vanish over all redshifts between the observer and the closest sources. This transformation was already illustrated in figure 1.12, where for a set of $\{\kappa_i\}$ and in the case of discrete sources, the transformation is possible for all planes

except the first two. The *nulled* selection functions are displayed with solid lines whereas the dashed lines of the same colours represent the corresponding regular selection functions.

Let me now reconsider Equation (4.23) for the BNT variable $\hat{\kappa}_i$,

$$\hat{\kappa}_i = \sum_j p_{ij}\kappa_j, \tag{4.27}$$

which only changes $w_i(r)$ into $\hat{w}_i(r)$ in expression (4.23). The main integral that appears in this expression can now be split into n_t different parts – in between the locations of the discrete source planes $\{r_i\}$. This nulled generating function then reads,

$$\hat{\phi}(\{\lambda_i\}) = \sum_{n=1}^{n_t} \int_{r_{n-1}}^{r_n} \mathrm{d}r\, \phi_{\mathrm{cyl/slope}} \left(\sum_{i=n}^{n+1} \lambda_i\, \hat{w}_i(r); r\theta_0 \right) \tag{4.28}$$

where $\lambda_{n_t+1} = 0$ and it is now clear that for each part, only two distinct subsequent λ variables of each nulled bin appear. In other words, the cumulant generating function takes the form

$$\hat{\phi}(\{\lambda_i\}) = \sum_{i=1}^{n_t-1} \tilde{\phi}_i(\lambda_i, \lambda_{i+1}) + \tilde{\phi}_{n_t}(\lambda_{n_t}). \tag{4.29}$$

Note that the functions $\tilde{\phi}_i$ are not cumulant generating functions on their own but are closely related to them. By identification and using the fact that $\hat{\phi}(\lambda_i) = \hat{\phi}(\{\lambda_j = \delta_{ij}\lambda_i\})$, it can then be shown that we have the following structure,

$$\hat{\phi}(\{\lambda_i\}) \;=\; \hat{\phi}(\lambda_1, \lambda_2) - \hat{\phi}(\lambda_2) + \hat{\phi}(\lambda_2, \lambda_3) - \cdots - \hat{\phi}(\lambda_{n_t-1}) + \hat{\phi}(\lambda_{n_t-1}, \lambda_{n_t}), \tag{4.30}$$

where $\hat{\phi}$ are now all genuine cumulant generating functions of the corresponding variables.

It thus remains that, with this choice of nulled variables, the joint cumulant generating function has a specific functional form: it is composed of a sum of functions that depend on two variables only, which implies that the full joint PDF can be obtained from bivariate CGFs (and hence PDFs) of neighbouring bins. I will thus now explore the consequences of such a feature on the computations and properties of the joint PDF.

One can first infer a number of general properties: it is clear that the two sets of variables $\{\hat{\kappa}_i\}_{i=1,\ldots,p-1}$ and $\{\hat{\kappa}_i\}_{i=p+1,\ldots,n_t}$ are correlated only through the variable $\hat{\kappa}_p$. In other words, once one marginalises over $\hat{\kappa}_p$, the two sets of variables are independent as there are no common structures that would contribute to both sets.

At the level of the cumulant generating function, one can notice that we have

$$\hat{\phi}\left(\{\lambda_i\}|_{\lambda_p=0}\right) = \hat{\phi}\left(\{\lambda_i\}_{i=1,\cdots,p-1}\right) + \hat{\phi}\left(\{\lambda_i\}_{i=p+1,\cdots,n_t}\right) \tag{4.31}$$

so that the cumulant generating function is split in two separate functions.

The application of equation (4.25) then readily shows that we have,

$$\int \mathrm{d}\hat{\kappa}_p\, \mathcal{P}\left(\{\hat{\kappa}_i\}_{i=1,\ldots,n_t}\right) = \mathcal{P}\left(\{\hat{\kappa}_i\}_{i=1,\ldots,p-1}\right)\, \mathcal{P}\left(\{\hat{\kappa}_i\}_{i=p+1,\ldots,n_t}\right). \tag{4.32}$$

This independence property for physically separated bins is specific to the nulled variables.

Finally, realistic tomographic cases are not made up of a collection of discrete source planes, but instead, source galaxies are split in redshift bins of finite width. The resulting nulling selection functions are then slightly more complex as illustrated in figure 1.13. One can indeed see here that the nulled profiles overlap with the first and second neighbours - and not only with the first. As a result, the functional form of the cumulant generating function is

$$\hat{\phi}\left(\{\lambda_i\}\right) = \tilde{\phi}\left(\lambda_1, \lambda_2, \lambda_3\right) + \cdots + \tilde{\phi}\left(\lambda_{n_t-2}, \lambda_{n_t-1}, \lambda_{n_t}\right) + \tilde{\phi}\left(\lambda_{n_t-1}, \lambda_{n_t}\right) + \tilde{\phi}\left(\lambda_{n_t}\right). \tag{4.33}$$

In this case, a property similar to equation (4.32) holds, except that an integration over two consecutive variables is required.

4.6.2 Impact of shape noise and intrinsic alignment on the BNT structure

The inclusion of shape noise in the joint nulled CGF leads to the addition of a term $\varphi_{\text{noise}}(\{\lambda_i\})$ (such that $\hat{\phi}_{\text{tot}} = \hat{\phi} + \varphi$) that follows the same functional form (4.33) whether realistic source bins or discrete source planes are considered, this because it automatically correlates lenses bins three by three with

$$\varphi_{\text{noise}}(\{\lambda_i\}) = \sum_{jii'} p_{ij} p_{i'j} \frac{\sigma^2_{\mathcal{S}}(j)}{2} \lambda_i \lambda_{i'} \tag{4.34}$$

where $\sigma^2_{\mathcal{S}}(j)$ is the amplitude of the shape noise in the original source bin j (and with respect to the shape noise, the bins are all independent) assuming it induces Gaussian noise. This amplitude for the Euclid experiment can for example be found in Deshpande et al. (2020) and was already used in the previous section. The only non-vanishing terms in (4.34) are those for which there is a j for which $j-2 \leq i \leq j$ and $j-2 \leq i' \leq j$ that implies that $|i-i'| \leq 2$ thus reproducing the form (4.33).

The shear-induced ellipticity of a galaxy can only be correlated to objects located within its radial selection function, as these correlations occur on smaller scales (typically below a hundred megaparsecs, and therefore not between distant redshift bins). Thus, in the case of nulling and similarly to shape noise, intrinsic alignments would also induce correlations that span over 3 nulled bins only, although with a priori a more complex dependence with λ since intrinsic alignments may not induce Gaussian noise. Still for the Euclid experiment, a modelling of the impact of intrinsic alignments on the lensing power spectrum and bi-spectra can be found in Deshpande et al. (2020), and used to obtain the second and third cumulants in the equivalent of equation (4.34) for the inclusion of galaxies alignments. However, there is no guarantee that such linear or quasi-linear models are valid in the regime we are probing here and more work is therefore necessary to model this effect properly, including small-scale baryonic effects (Codis et al., 2015).

This thus implies that the correlation structure I presented for nulled variables is still valid when such effects are taken into account. Note however that corrective terms coming from reduced shear, magnification bias or post-Born corrections formally break this structure, although weakly, as they change the functional form of the convergence and introduce couplings between lenses, which are not considered independent anymore. I have already presented in this thesis several computations for these effects, especially in chapter 1. Finally, note that one obvious but strong limitation of this procedure is the precision to which

the redshifts of sources inside tomographic bins can be determined (photometric redshift errors). The quantification of some of these effects on the nulling strategy will be performed in future works in preparation.

4.6.3 Numerical complexity of the joint PDF

Let me now explore the resulting numerical complexity of the computation of the joint PDF of nulled maps, $\{\hat{\kappa}_i\}$. Note that this is not just a numerical trick to reduce the computing time, this is the simpler form the joint PDF can take using the correlation structures that we have. In particular, equation (4.24) for nulled bins can as a result be written as some function of joint PDFs between 2 or 3 bins but not in a simpler form than the one we present here, see for example appendix A.5 for an exposition of this fact.

Let me start this analysis in the case of discrete source planes, that is, taking advantage of the functional form (4.29) for the cumulant generating function and thus assuming that this generating function can be written as a sum of functions of two consecutive variables.

If we assume that the number of operations for a given integral over λ_i is N, then, and in the absence of factorisation properties, the a priori complexity for the computation of the joint κ-PDF, as given by equation 4.24, is N^{n_t}. The purpose of this subsection is then to show how this complexity, and hence the computational time, can be reduced when changing variables from κ to $\hat{\kappa}$. To reach this goal, let us define the function $\zeta_2(n_t)$, supposedly an integer, that gives an estimate of the number of operations to be done as $N^{\zeta_2(n_t)}$. Similarly, I also define a function $\zeta_3(n_t)$ for the functional form (4.33) of the cumulant generating function corresponding not to successive localised source planes but wide redshift bins.

Let me start with a simple case. For a joint analysis of three source planes, the cumulant generating function takes the form,

$$\hat{\phi}(\lambda_1,\lambda_2,\lambda_3)=\hat{\phi}(\lambda_1,\lambda_2)-\hat{\phi}(\lambda_2)+\hat{\phi}(\lambda_2,\lambda_3) \tag{4.35}$$

so that we can write,

$$\mathcal{P}(\hat{\kappa}_1,\hat{\kappa}_2,\hat{\kappa}_3)=\int\frac{\mathrm{d}\lambda_2}{2\pi\,\mathrm{i}}\,e^{-\lambda_2\hat{\kappa}_2+\hat{\phi}_{\hat{\kappa}_1}(\lambda_2)-\hat{\phi}(\lambda_2)+\hat{\phi}_{\hat{\kappa}_3}(\lambda_2)} \tag{4.36}$$

where

$$e^{\hat{\phi}_{\kappa_i}(\lambda_j)}=\int\frac{\mathrm{d}\lambda_i}{2\pi\,\mathrm{i}}\,\exp(-\lambda_i\hat{\kappa}_i+\hat{\phi}(\lambda_i,\lambda_j)). \tag{4.37}$$

The computation of each function $e^{\hat{\phi}_{\kappa_i}(\lambda_j)}$ simply scales like N^1. This thus implies that the estimated number of operations for the computation of expression (4.36) is $N\times(N+N)$ which scales like N^2 so that $\zeta_2(3)=2$.

Let me now explore the general case and define the quantity

$$e^{\hat{\phi}_{\kappa_p}(\lambda_{p-1},\lambda_{p+1})}=\int\frac{\mathrm{d}\lambda_p}{2\pi\,\mathrm{i}}\,\exp\left[-\lambda_p\hat{\kappa}_p+\hat{\phi}(\lambda_{p-1},\lambda_p)-\hat{\phi}(\lambda_p)+\hat{\phi}(\lambda_p,\lambda_{p+1})\right]. \tag{4.38}$$

The number of operations for the computation of this function is N. The idea is now, as exemplified in the simple 3-bin case, to decimate the variables, 1 out of 2, as illustrated on figure 4.12.

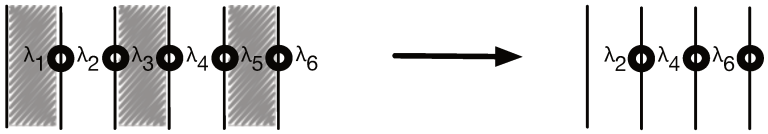


Figure 4.12: Schematic representation of the decimation procedure for the case of discrete source planes. The generating function is a sum of functions of adjacent variables. Each of these terms is depicted with a dark circle. The integrations over 1 variable out of 2, (grey areas) can then be made independently. The resulting structure is again a sum of functions of adjacent variables and one can thus apply this decimation strategy recursively. This figure was published in **Barthelemy** et al. (2021a).

Let me first assume that the number of planes, n_t, is an odd number, $n_t = 2m + 1$. From the structure of the cumulant generating function, and the previous definition, follows one of the main results of this subsection

$$\mathcal{P}(\{\hat{\kappa}_i\}) = \int \prod_{i=1}^{m} \frac{\mathrm{d}\lambda_{2i}}{2\pi\,\mathrm{i}}\, e^{-\sum_{i=1}^{m}\lambda_{2i}\hat{\kappa}_{2i}} \exp\left[\sum_{i=0}^{m} \hat{\phi}_{\kappa_{2i+1}}(\lambda_{2i},\lambda_{2(i+1)}) - \hat{\phi}(\lambda_{2i})\right] \tag{4.39}$$

with $\lambda_0 = \lambda_{n_t+1} = 0$. The number of operations required to integrate the odd variables is of the order of $(m+1)N$. What is then remarkable is that the structure of the resulting m-inverse Laplace transform is the same as before: the effective cumulant generating function is a sum of function of two adjacent variables only. As a result the expected number of operations to be performed is $(m+1)N \times N^{\zeta_2(m)}$. To a logarithmic correction I will ignore in the following we thus have $\zeta_2(2m+1) = 1 + \zeta_2(m)$. Let me finish this evaluation by noting that there is no gain in reduced numerical complexity going from $2m+1$ to $2m$ source planes. We thus have $\zeta_2(2m) = \zeta_2(2m+1) = 1 + \zeta_2(m)$.

Applying this simple rule to the first few integers, I find,

$$\begin{aligned} \zeta_2(1) &= 1 \\ \zeta_2(2) = \zeta_2(3) &= 2 \\ \zeta_2(4) = \cdots = \zeta_2(7) &= 3 \\ \zeta_2(8) = \cdots = \zeta_2(15) &= 4. \end{aligned} \tag{4.40}$$

Finally note again that the reason we need to go through this somewhat complicated numerical integration scheme comes from the fact that there is no simpler form for the joint $\hat{\kappa}$-PDF in terms of bivariate PDFs of neighbouring bins as is the case for the CGFs. To make this explicit, I show in appendix A.5 how the PDF of 3 neighbouring bins is expressed explicitly as a function of bivariate PDFs.

I now move to the case of equation (4.33). We can use the same trick and decimate the variables but I have however identified here two possible strategies which are depicted in figure 4.13.

The first one is to decimate the variables as before: 1 variables out of 2. The difference here is that this first set of variables are not uncorrelated, they are actually correlated 2 by 2. To be more precise let us again assume that $n_t = 2m + 1$ and the complexity of this

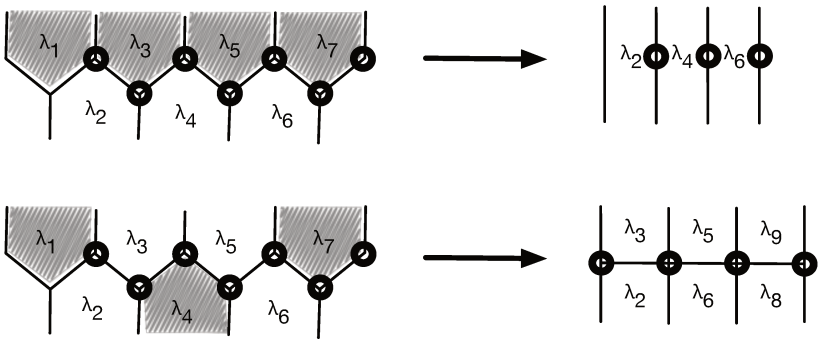


Figure 4.13: Schematic representation of the decimation procedures for the case of extended source distributions. The generating function is a sum of functions of 3 adjacent variables. Each of these terms are depicted with a dark circle. In the top panel the proposed scheme is to first consider the integration of 1 variable out of 2 (shaded odd variables). These are correlated 2 by 2 and we thus apply the first step depicted in figure 4.12. We then consider the even variables which are also correlated 2 by 2. This allows to once again perform the first step of figure 4.12 and we then go back to the odd variables and apply this scheme recursively. In the bottom panel the procedure is to integrate 1 variable out of 3 where each of those integrations can be performed independently. This leads to an effective cumulant generating function for which variables are correlated via adjacent pairs which we again know how to treat. This figure was published in **Barthelemy** et al. (2021a).

operation is thus $N^{\zeta_2(m+1)}$ defined in the previous section. We are then left with m variables that are open again correlated 2 by 2 and therefore have

$$\zeta_3(2m+1) = \zeta_2(m+1) + \zeta_2(m) \tag{4.41}$$

and similarly

$$\zeta_3(2m) = 2\zeta_2(m). \tag{4.42}$$

The second strategy (the resulting complexity being denoted with a new function ζ_3') is to decimate 1 variable out of 3, so that the first set of variables are uncorrelated. For this operation, the number of operations scales like N. To be more precise, let me assume $n_t = 3m+1$. After decimation we are left with $2m$ variables. These variables then follow a specific structure. They form pairs that are correlated to their nearest neighbours, which is a situation we know how to handle. Each pair introduces an N^2 factor so that we eventually have,

$$\zeta_3'(3m+1) = 1 + 2\zeta_2(m). \tag{4.43}$$

With such a strategy we have $\zeta_3'(3m-1) = \zeta_3'(3m) = \zeta_3(3m+1)$ (except for $\zeta_3'(1)$ and $\zeta_3'(2)$).

It turns out that these two strategies have competing performances, as illustrated in table 4.1, and both drastically reduce the complexity of a naive implementation scaling like N^{n_t}.

4.6.4 Practical implementation

Let me now move to the actual implementation of the decimation strategy presented in the last section. I begin with a brief reminder of the computation of ϕ_{cyl} and ϕ_{slope} and their analytical continuation in the complex plane. The computation of $\hat{\phi}(\{\lambda_i\})$ is then performed on a fixed grid of imaginary values $\{\lambda_i\}$, and I finally illustrate the procedure described in the previous section with several plots.

n_t	$\zeta_2(n_t)$	$\zeta_3(n_t)$	$\zeta_3'(n_t)$
1	1	1	1
2	2	2	2
3	2	3	3
4	3	4	3
5	3	4	5
6	3	4	5
7	3	5	5
8	4	6	5
9	4	6	5
10	4	6	5
11	4	6	7
12	4	6	7
13	4	6	7
14	4	6	7
15	4	7	7

Table 4.1: Numerical complexity $N^{\zeta(n_t)}$ for the calculation of the joint κ-PDF in case of discrete source planes, $\zeta_2(n_t)$, and extended source planes, $\zeta_3(n_t)$ and $\zeta_3'(n_t)$, as a function of the number of source planes. The latter two results correspond to two different strategies as described in the main text. This table was published in **Barthelemy** et al. (2021a).

Computation of $\hat{\phi}(\{\lambda_i\})$: The computation of this quantity has already been described previously both for the convergence and the aperture mass using a formalism inspired by large deviation theory, though some equations were already known but not interpreted in this framework, as in Bernardeau & Valageas (2000).

In this context, the joint cumulant generating function of the density filtered in concentric long cylinders of transverse sizes $\{R_i\}$ at redshift z is given by

$$\phi_{\{\delta_i\}}(\{\lambda_i\}) = \sum_i \lambda_i \delta_i - \Psi_{\{\delta_i\}}(\{\delta_i\}), \tag{4.44}$$

where $\{\delta_i\}$ are functions of $\{\lambda_i\}$ through the stationary conditions

$$\lambda_k = \frac{\partial \Psi_{\{\delta_i\}}(\{\delta_i\})}{\partial \delta_k}, \quad \forall k. \tag{4.45}$$

The rate function $\Psi_{\{\delta_i\}}(\{\delta_i\})$ is given by

$$\Psi_{\{\delta_i\}}(\{\delta_i\}) = \frac{1}{2} \sum_{k,j} \Xi_{kj} \mathcal{C}^{-1}(\delta_k) \mathcal{C}^{-1}(\delta_j), \tag{4.46}$$

where Ξ_{kj} is the inverse of the non-linear covariance matrix between cylinders of transverse radii $\{R_i\,(1+\delta_i)^{1/2}\}$, and the inverse of the cylindrical collapse dynamics $\mathcal{C}$ can be approximately written as

$$\mathcal{C}^{-1}(\delta_k) = \nu - \nu(1+\delta_k)^{-1/\nu}, \quad \nu = 1.4. \tag{4.47}$$

Then, from equation (4.44) and the generic properties of cumulant generating functions, one can finally define

$$\phi_{\rm cyl}(\lambda) = \phi_\delta(\lambda) \quad \text{and} \tag{4.48}$$
$$\phi_{\rm slope}(\lambda) = \phi_{\delta_1,\delta_2}(-\lambda,\lambda). \tag{4.49}$$

The previous set of equations thus allows one to numerically compute $\phi_{\rm cyl}(\lambda)$ and $\phi_{\rm slope}(\lambda)$ for any real value of λ. Unfortunately, since the computation of the PDF of projected densities requires an integration in the complex plane, we need a prescription for the analytic continuation of $\phi_{\rm cyl/slope}$. A good strategy to keep the analytical properties of the construction (4.44) consists in fitting an effective $\mathcal{C}$ function, rewriting equation (4.44) for $\phi_{\rm cyl/slope}$ as

$$\phi_{\rm cyl/slope}(\lambda) = \lambda\mathcal{C}(\tau_{\rm eff}) - \frac{1}{2}\tau_{\rm eff}^2, \tag{4.50}$$

with the effective stationary condition then written as

$$\lambda = \frac{\rm d}{{\rm d}\mathcal{C}}\frac{\tau_{\rm eff}^2}{2} = \tau_{\rm eff}\left(\frac{{\rm d}\mathcal{C}(\tau_{\rm eff})}{{\rm d}\tau_{\rm eff}}\right)^{-1}. \tag{4.51}$$

Note then that since

$$\frac{{\rm d}\phi_{\rm cyl/slope}(\lambda)}{{\rm d}\lambda} = \mathcal{C}(\tau_{\rm eff}), \tag{4.52}$$

and hence

$$\frac{1}{2}\tau_{\rm eff}^2 = \lambda\frac{{\rm d}\phi_{\rm cyl/slope}(\lambda)}{{\rm d}\lambda} - \phi_{\rm cyl/slope}(\lambda), \tag{4.53}$$

one can thus fit both the values of $\tau_{\rm eff}$ and $\mathcal{C}(\tau_{\rm eff})$ from the generating function computed for real values of λ. An expansion in series of $\mathcal{C}$,

$$\mathcal{C}(\tau_{\rm eff}) = \sum_{k=0}^{n}\frac{\mu_k}{k!}\tau_{\rm eff}^k, \tag{4.54}$$

typically stopping at $n = 5$, allows one to fit the μ_k coefficients, which are closely related to the cumulants of $\phi_{\rm cyl/slope}$. This construction finally enables one to successfully compute $\phi_{\rm cyl/slope}$ and thus the joint cumulant generating function of projected densities $\hat{\phi}(\{\lambda_i\})$ in equation (4.28) for any tuples of complex $\{\lambda_i\}$.

RESULTS: I implement the decimation strategy coupled with the computation of $\hat{\phi}(\{\lambda_i\})$ of the last paragraph. For simplicity and visualisation purposes I restrict myself to the computation of the joint convergence between 3 successive nulled bins taken from figure 1.12, the 5th, 6th and 7th, and which are respectively the linear combination of source redshifts located at $z_s = 0.62 - 0.73 - 0.85$, $0.73 - 0.85 - 0.96$, and $0.85 - 0.96 - 1.1$. Thus denoting by respectively $\hat{\kappa}_5$, $\hat{\kappa}_6$ and $\hat{\kappa}_7$ the nulled convergence in each of those bins, and noticing that their joint PDF is given by equation (4.36), we obtain the results given in figure 4.14 for 1 and 2D marginals and for the full 3D PDF. As expected from the formalism, the exponential decay of the high-$\hat{\kappa}$ tails are visible on all these plots and the highly skewed (asymmetric) PDFs hints towards the fact that the information content of the projected densities fields

is not entirely probed by its 2-point correlation function which is sufficient for Gaussian random fields only.

Figure 4.14 also displays the shape of the correlation between $\hat{\kappa}_5$ and $\hat{\kappa}_6$ seen as the residual between the 2D PDF and the product of the two 1D PDFs. This residual would be exactly equal to 1 if the two nulled variables were independent. Note that this modulation of the independent-variables case is quite important and understandable as follows: $\hat{\kappa}_5$ and $\hat{\kappa}_6$ result from the overlapping contributions of lenses along the line of sight. As a consequence, if $\hat{\kappa}_5$ were to take an improbable very high (resp. low) value, then the modulation would have to take that into account by raising the probability of finding a very high (resp. low) value for $\hat{\kappa}_6$ compared to the independent-variables case.

I also show on figure 4.15 the correlation induced between $\hat{\kappa}_5$ and $\hat{\kappa}_7$ if a constraint on $\hat{\kappa}_6$, here $\hat{\kappa}_6 > 0$, is imposed. In the absence of a constraint, these two variables are independent (since their nulled lensing kernels do not overlap) but the constraint breaks this property. In this case, one needs to compute the full 3-variable PDF before integrating over the constraint. The residual between the 2D constrained PDF and the product of the two 1D constrained PDFs is relatively small compared to the previous case (typically between 50 and 150%) and points out an over-probability, compared to the independent assumption, for $\hat{\kappa}_5$ and $\hat{\kappa}_7$ to have opposite rare values while same sign rare events are less likely. Note that since $\hat{\kappa}_6$ has its own distribution the constraint $\hat{\kappa}_6 > 0$ is more likely to give values of $\hat{\kappa}_6$ close to zero, and thus the constraint imposes for example that for a large positive value of $\hat{\kappa}_5$, $\hat{\kappa}_7$ should compensate in the opposite direction to realise it. Hence the $\hat{\kappa}_7$ distribution would be shifted towards more negative values compared to its marginal distribution. For $\hat{\kappa}_6 = 0$, $\hat{\kappa}_5$ and $\hat{\kappa}_7$ would be exactly anti-correlated but this is not exactly the case here as seen from the fact that figure 4.15 is not exactly symmetric with respect to the $\hat{\kappa}_5 = \hat{\kappa}_7$ axis.

4.6.5 Prospects for the nulling strategy on the PDF

Regardless of the model – (semi-)analytical or derived from numerical simulations – used to infer the statistical properties of the underlying matter field, the nulling formalism that I presented ought to be implemented in the context of cosmological analysis of weak lensing surveys. Indeed, while it preserves the information content of the observables, it allows one to better identify the origin of this information, whether it is in terms of physical scales, redshift, or combinations of those. In this context, my results show that the actual numerical derivation of joint PDFs is tractable even if one wants to exploit a large number of bins in order to be as precise as possible while preserving the amount of information available. Also note that the method presented in this section can also be used to obtain the joint PDF of the regular fields via a simple change of variables once the joint PDF of nulled bins has been computed.

For the specific case of non Gaussian observables such as the (joint) PDF of projected densities, it was shown that, for the desired filtering angular scales, shape noise in individual nulled bins makes the detection of non-Gaussian features virtually impossible in a single nulled map. One of the reasons of this failure is that each nulled map collects noise from 3 of the original maps it is built from. However, the exploitation of joint PDFs now offers a potential solution to this issue as more useful signal can then be combined together while the shape noise does not increase in the same proportion since it is shared between nulled

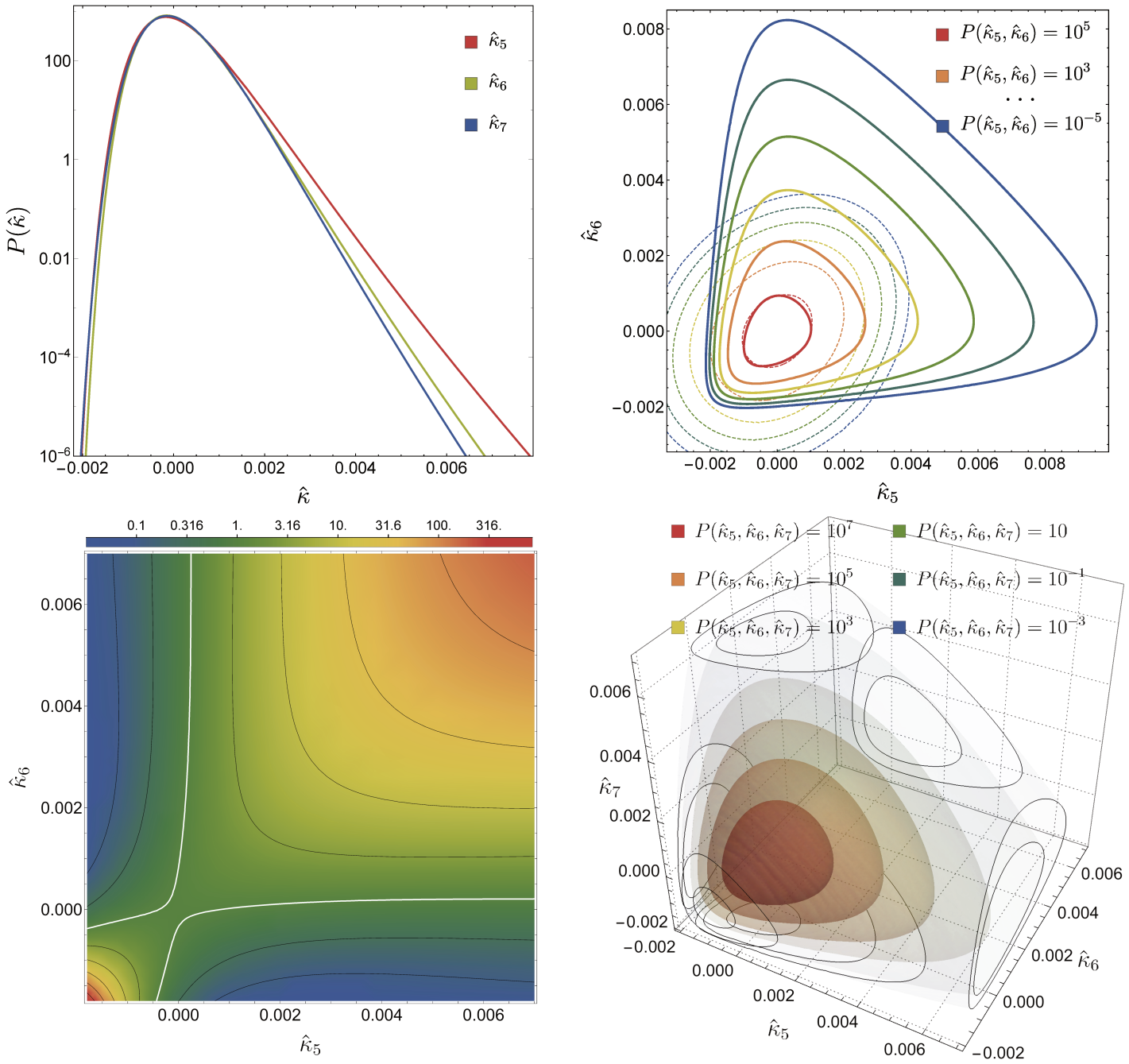


Figure 4.14: (joint) PDFs between the 3 nulled bins filtered by a top-hat window function of radius 15 arcmin. They are respectively the linear combination of source redshifts located at $z_s = 0.62 - 0.73 - 0.85$, $0.73 - 0.85 - 0.96$, and $0.85 - 0.96 - 1.1$. Top left panel: One-point $\hat{\kappa}$-PDF within the 3 nulled bins. Top right panel: Iso-probability density contours of the joint $\hat{\kappa}$-PDF between the 2 first nulled bins. The nulled fields are filtered by a top-hat window function of radius 15 arcmin. The dashed thin lines correspond to a 2-dimensional normal distribution with the same covariance. Bottom left panel: Density plot of $P(\hat{\kappa}_5, \hat{\kappa}_6)/P(\hat{\kappa}_5)/P(\hat{\kappa}_6)$ to explicit the shape of the correlation between $\hat{\kappa}_5$ and $\hat{\kappa}_6$. The black lines correspond to iso-contours of values the ticks of the colour-bar and the white line is for the tick equal to 1. They serve as guide for the eye. Bottom right panel: Joint $\hat{\kappa}$-PDF between the 3 nulled bins. This figure was published in **Barthelemy** et al. (2021a).

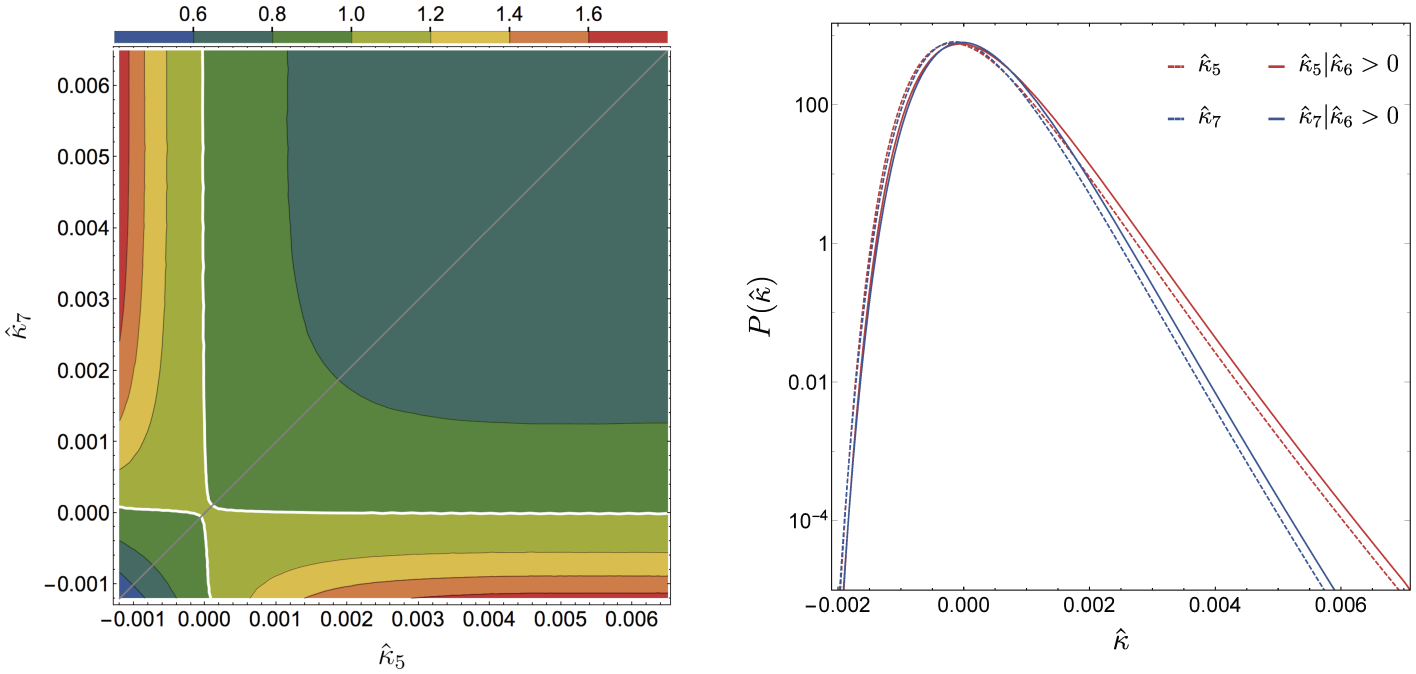
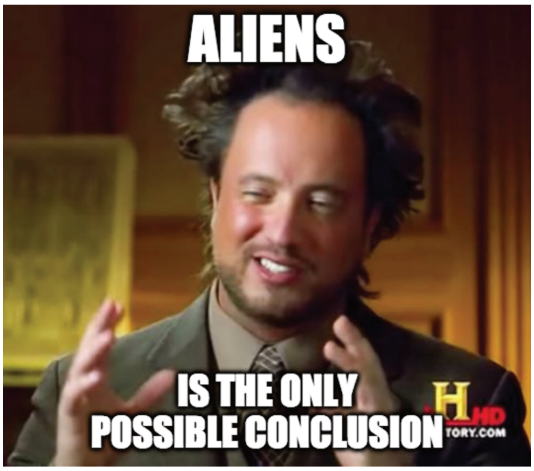


Figure 4.15: Left-hand panel: Contour plot of $P(\hat{\kappa}_5, \hat{\kappa}_7 \,|\, \hat{\kappa}_6{>}0)/P(\hat{\kappa}_5 \,|\, \hat{\kappa}_6{>}0)/P(\hat{\kappa}_7 \,|\, \hat{\kappa}_6{>}0)$ to explicit the shape of the correlation between $\hat{\kappa}_5$ and $\hat{\kappa}_7$ if a constraint on $\hat{\kappa}_6$ is imposed. The white line is for the tick equal to 1. The nulled fields are filtered by a top-hat window function of radius 15 arcmin. Right-hand panel: The solid lines represent $P(\hat{\kappa}_5 \,|\, \hat{\kappa}_6{>}0)$ and $P(\hat{\kappa}_7 \,|\, \hat{\kappa}_6{>}0)$ while the dashed lines show the same PDFs as the top left panel of figure 4.14, that is without any constraint. Together the two panels give $P(\hat{\kappa}_5, \hat{\kappa}_7 \,|\, \hat{\kappa}_6{>}0)$. This figure was published in **Barthelemy** et al. (2021a).

bins. This thus makes the detection of sought-after non-Gaussian features more likely.

At this stage I can sketch a possible strategy: the key to a successful exploitation of the data is our ability to select the physical scales that are contributing the most to the signal and reject the smallest scales, in practice the nearby lenses, for which we miss reliable modelling and which incidentally are likely to be strongly affected by large super-sample covariance effects. The idea would then be to exploit joined nulled bins, built out of a large set of n_t bins, but a priori ignoring the 2 closest for which the physical scales are still strongly mingled. Hopefully the $n_t - 2$ farthest bins would be able to provide us with significant detection of the non-linear couplings in a regime where they are well understood. And if not, it would point to the fact that the non-Gaussian information that we observe in regular bins is coming from poorly understood or modelled non-linear scales, which would hence suggest a need for extreme caution when using them to extract information. This is quite unlikely as we were able to extract usable information from the convergence PDF even staying in a regime where our theoretical approach remains valid. In any case, the nulling formalism and the joint PDF I presented here could play a key role in answering these questions. Such studies are however beyond the work I did over the course of my thesis and are left for future projects.



Conclusion

To conclude my manuscript and to ease the workload of readers that might want to first check the conclusion before anything else, let me start by a summary of the content of each chapter. I hope that the unity and coherence between those is transparent that way: I build the theoretical formalism for the statistics of weak-lensing from that of the density field in chapter 1, introduce how to compute the one-point statistics of the density through the large deviations formalism in chapter 2, and then build the convergence and aperture mass one-point statistics by joining the results from the two first chapters. I then move on to a more general conclusion, first by introducing immediate projects that could be carried out in continuation of the work presented here, and then by trying to provide a more general picture of the utility of this work in the future.

Summary of each chapter

Chapter 1: This first chapter aims at giving an introduction to the basic formalism at the heart of the theoretical interpretation of cosmic-shear experiments. I derive an evolution equation for the deformation matrix of thin light beams along the line of sight (Sachs equation) from basic general relativistic considerations, and solve it in a perturbative way up to next to leading order for a scalar-perturbed ΛCDM universe. This then allows me to present the usual weak-lensing formalism and the approximations that are made (Born approximation and independent lenses), as well as standard additional observational considerations that need to be taken into account (shape noise, intrinsic alignments, baryonic feedback and redshift space distortions). As an exercise and to demonstrate the computation of high-order corrections to the standard formalism, I then present what is known as post-Born corrections to the convergence field and compute their main observational consequences on the one-point statistics of the convergence field by computing the post-Born-induced third cumulant. This should not be taken as a general calculation on the importance of post-Born effects as the skewness is a really specific and rather unusual observable, but rather as an example of the type of computations that need to be performed to take into account post-Born corrections. Moreover this correction on the skewness helps with describing later on in chapter 3 how to include these corrections on the PDF of the convergence field. I then make a small detour towards the use, some limitations and complementarity with theoretical computations of numerical simulations in our field. This serves as a good reminder of the need for theoretical developments of the observables in the large scale structure community. I finally introduce

the BNT transform, a very important observational strategy for the following parts of this thesis, mostly chapters 3 and 4 on the convergence and aperture mass. This transform, which exploits some sort of hidden symmetry in the weak-lensing formalism, allows us to build new observables from linear combinations of observed maps along the line of sight which possess several desirable properties. i) The new nulled observables are mostly uncorrelated – only correlated 2 by 2 – ii) the translation of angular scales (ℓ) to physical ones (k) is defined more clearly and the range of k probed by one l can be narrowed almost at will allowing for the use of quasi-linear theoretical models, and iii) there is absolutely no loss in information whatsoever going from the regular variables to the nulled ones (it is an invertible linear application). The BNT transform is introduced both for source planes and realistic source distributions, and I present some calculation on the impact of some approximation and observational systematics that tend to break the nulling property by introducing additional couplings between the lenses. Those impact are found to be very small.

Chapter 2: This second chapter serves as an introduction to the mathematical theory of large deviations. After a quick historical introduction I start by giving a typical example of a random variable, the sample mean of independent and identically distributed random variables, satisfying a large deviation principle. In simple terms, this principles states that when some natural parameter n of the problem varies and tends to infinity, the probability distribution function of the considered random variable X decays exponentially as $\exp(-n\psi(x))$ with ψ being the rate function. As such it can look like an extension of the central limit theorem in which case ψ is just quadratic in x around its minimum. This serves as a starting point to introduce key results of large deviation theory such as the Gärtner-Ellis and Varadhan's theorems, and the contraction principle. The first two roughly allow us to link the cumulants of a random field and its rate function, while the latter allows us to compute the rate function of a random variable B provided that it is linked through a continuous mapping to another random variable A which admits a large deviation principle and whose rate function is known. After a small detour towards statistical mechanics to illustrate how large deviation theory can be seen as the rigorous mathematical language behind, I then show how the cosmic matter field filtered in a spherically symmetric filter can be shown to satisfy a large deviation principle, highlight that its driving parameter is the inverse of the variance of the field, $1/n = \sigma_\delta^2$, and calculate its rate function. Following this is a discussion on how to extend the results given by large deviation theory outside the large deviations regime (that is for finite values of σ_δ^2) and the formal link that can be drawn between the results we thus obtain on the statistics of the density field and Newtonian Eulerian perturbation theory at leading order. Finally, I present how through the use of the contraction principle one can establish a large deviation principle for any monotonic function of the cosmic matter density field, functions which, if carefully chosen, might lead to more accurate results for the statistics of the density field once we extend the results beyond the large deviations regime. I conclude with a literature survey on how those results have allowed the accurate theoretical modelling of many observables based on the PDF of the matter density field. The original results of my thesis then consist in extending the formalism presented in this chapter to weak-lensing observables, and more precisely the PDFs of the convergence and aperture mass.

Chapter 3: This third chapter starts with a literature review on the state of the art in cosmological parameters constraints in cosmic shear experiments. The main message is as follows. First of all the most used tools are a combination of different two-point statistics (3x2 points analysis) from which we derive most of our constraints. However we know that two-point statistics contain only complete statistical information for Gaussian random fields which is only the case for the density field in the linear regime of structure formation, neglecting non-Gaussianities that may come from an earlier inflationary phase. There is thus also a strong interest in the literature for so-called higher-order statistics such as higher-order correlation functions (bispectrum etc...), peak counts, Minkowski functionals or PDFs. These observables are all found to provide complementary constraints on cosmological parameters to the power spectra, but they are often modelled purely from numerical simulations (with a few exceptions). For the specific case of the PDF, there exist several phenomenological models for the PDF that notably rely on the approximation that it is log-normal, or on the halo model. More physically motivated models for the lensing PDF were built in the early 2000's, for example relying on hierarchical models. Nevertheless, there is room – and necessity because of uncertainties coming from purely numerical analysis – for theorists to build accurate theoretical models for the lensing PDF. That is what I do in this chapter. I thus start by deriving the projection formula that allows us to link the cumulant generating function of the convergence to that for infinite cylinders at different redshifts, and whose radii are chosen so as to reproduce the light cone from the observer to the source. After discussing how to treat some technical details on the implementation of large deviation theory to the convergence, which appears not to satisfy a large deviation principle itself, I perform a comparison between our theoretical cumulant generating functions of the convergence – itself an observable at the same level than the PDF – to the one measured in some state-of-the-art full-sky weak-lensing numerical simulation. Again after discussion on some details for the technical implementation of the PDF, I also compare the theoretical PDF to the one measured in the numerical simulation. Be it for the PDF or CGF, it turns out that though our prediction is accurate for the most probable events, we lack a good description for the large tails, the rare events, of the convergence field. We advocate that this is due to scale-mixing of linear and quasi-linear physical scales with very small, non-linear ones non accessible to our theoretical formalism limited to quasi-linear physical scales. This scale-mixing happens inevitably for projected quantities such as lensing observables and I thus introduce the use of the BNT transform described in the first chapter to get accurate predictions for the convergence PDF/CGF. I demonstrate that this allows us to obtain very precise predictions, accurate even in the far tails of the PDF, for angular scales and source redshifts quite inaccessible to theoretical descriptions without the BNT transform. Finally, this chapter also presents a full Fisher forecast that is made with derivatives of the PDF with respect to cosmological parameters coming from our theoretical formalism for the convergence PDF. I show that these derivatives are perfectly accurate by comparing them to numerical simulations, and I advocate that these are even better because numerical noise in the derivatives tend to be taken as specific signals which thus unrealistically improve constraints on cosmological parameters. The covariance matrices for the PDF are them estimated using fast log-normal realisations of the convergence field which are shown to be sufficient to obtain meaningful constraints. Taking into account shape noise, considering only one source redshift ($z_s = 2$) and two smoothing scales ($\theta = 7$ and 10 arcmin) I arrive at

the improved constraints summarised in table 3.4 when considering the Planck experiment, the PDF and the convergence two-point correlation function. Overall it is clear that there is a large benefit in considering the lensing PDF for cosmological parameters constraints, it is relatively easy to measure, does not produce a huge data-vector, contains non-negligible additional information, and can be reliably theoretically predicted which spoils the need for expensive and numerous numerical simulations for each cosmological model.

Chapter 4: This final chapter is a direct thematic continuation of the third one. There I move from the theoretical description of the PDF of the BNT convergence to the BNT aperture mass. I start by formally defining the aperture mass as the difference between the convergence field filtered at two different scales, I adapt the projection formula from the third chapter to this new configuration and again discuss the technical details around the implementation of the CGF and the PDF. This allows us to once again perform some comparisons of the formalism of the aperture map PDF to results coming from the same numerical simulation as in chapter 3. There I find that there is once again a regime of perfect agreement between the theory and the simulated PDF. However, it turns out that the aperture mass as measured in the simulation from a difference of the κ field filtered at two different scales thus suffers from very little numerical artefacts on the simulated convergence that are very much visible when looking at the aperture mass. As a consequence, it turns out that I measure some contradictions in the simulation itself for the case of the aperture mass which thus prevent me from doing a more careful assessment of the validity regime of our theoretical formalism as I did for the convergence in chapter 3. Nevertheless, the formalism is there and can be tested in more detail if possible/needed at a later time, or even used to test some details of future numerical simulations. One known issue with the BNT transform that I explicitly demonstrate in this chapter is the fact that the noise, for example shape noise, on a single BNT observable (here convergence or aperture mass) is actually much more important that for a usual one. The trick resides in the fact that the level of noise is actually equivalent for the BNT and regular observables only when one considers the full line of sight (all source redshifts, up to the survey limit) in a tomographic analysis. As such, for our formalism with large deviation theory and the BNT transform to be applicable in a real context, we actually need an efficient way to predict the full joint PDF of the BNT observables along the line of sight. This is thus exactly what I do in the last part of this final chapter. Taking advantage of the specific correlation structure of the BNT observables along the line of sight, I introduce a numerical strategy to perform the numerous nested complex integrals that lead to the joint PDF. The presented strategy proves to be particularly efficient compared to the naive implementation and is only possible making use of the BNT variables. Note however that the full joint PDF of BNT variables along the line of sight allows us through a simple change of variable to obtain the full-joint PDF of the regular variables. I conclude this chapter with several prospects on the implementation of this strategy with respect to shape noise. New results along those lines are unfortunately not present in this thesis but are part of the projects that I am now working on.

General conclusion and prospects

Generally speaking, I would describe the general research program in which the work presented in this thesis takes place as: "building new or known statistical observables of the large scale structure from strongly physically motivated theoretical formalisms, and trying to push their modelling as close as possible to real-life measurement conditions". In this sense, the formalism presented in this thesis for the (joint) PDF of the BNT convergence/aperture mass is very much in this line: We started from the basic theoretical consideration "The cosmic density field filtered in spherical windows satisfies a large deviation principle" and ended up with the built PDF of a direct observable, the aperture mass. Scale-mixing was an issue in providing meaningful predictions, and thus we introduced the BNT transform. In turn studying only one BNT aperture mass field proved to be an issue for noise considerations, so we introduced tractable means to compute the full joint PDF.

As such, immediate prospects surrounding the research presented in this thesis would have to follow the same philosophy. I can divide those future projects into several categories, first there are projects that build on the type of results that I have presented with large deviation theory and that are merely extensions of it, some concern typically the use of the BNT transform in our formalism, and some are simply new ideas that go back a little and try to lift some of the assumptions that we made and try new things. Let me start by projects in direct continuation of the present work with large deviation theory. One thing that may come to mind is the extension of the whole formalism presented for the matter and thus convergence PDF to theories of gravity other than general relativity and different models of dark energy. Indeed it is interesting that our model is accurate for a whole range of ΛCDM cosmologies making use of an EdS collapse and it would be nice to check what needs to be modified for those models. This project just appeared on the arXiv at the time of writing of those lines (Cataneo et al., 2021). In the specific context of the PDF of weak-lensing observables, one observational systematic from section 1.5.4 whose impact was not evaluated is intrinsic alignments which should be done at some point. On the fact that because subtle non-linear effects seem to matter for the statistics of the aperture mass, results from the spherical collapse dynamics may need to be refined. This could be done in the same way that we include post-Born corrections, that is by computing the one-loop correction to the skewness of the density field, implementing it by a change in the fit of the spherical collapse ν parameter, or this could also be done implementing the log mapping of section 2.5.3 in the line of sight integrations that lead to κ and $M_{\rm ap}$. This might at the same time allow us to enlarge the domain of applicability of the model towards possibly smaller scales and redshifts. The formalism in its current state is also available to compute joint statistics between clustering and lensing, or lensing at small redshifts and at the last scattering surface, although for the halo and galaxy PDF, we would still need to explore how to translate everything in redshift rather than real space. This is not trivial because as mentioned in section 1.5.4, RSDs are well taken into account only in the linear regime (Kaiser effect) which thus prevents us from getting an accurate PDF at the scales at which we are typically looking at (~ 10 Mpc/h). This comes from the fact that though we can get accurate modelling in the quasi-linear regime for both the density and the velocity fields through the spherical collapse dynamics, the joint modelling of the two, or said differently the velocity dispersion for each value of the density, is not well known. For the issue of shot

noise and halo/galaxy bias for the PDF, Friedrich et al. (2021) corresponds to the state of the art and is however proven to work extremely well.

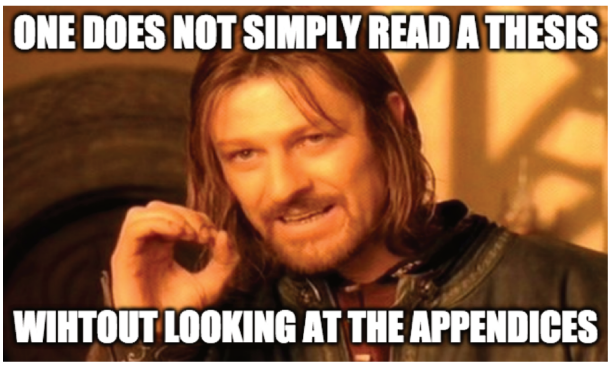
For the specific case of the application of BNT transform to model our lensing PDFs, several directions should be taken. A real study with the help of hydro or modified N-body simulations to account for the impact of baryons should be carried out to get a precise idea of the redshift bins one should cut to extract information from the scales for which we have reliable and understood theoretical models. This type of study should also at some point evaluate the impact of photometric redshift errors, redshift space distortions and reduced shear corrections on the exactness of the BNT procedure. Taking into account the previous considerations, we should also explore the viability of the BNT transform as a cosmological probe in itself, for example checking if the cross-correlation between the shear and galaxy densities from un-correlated nulled redshift bins is effectively zero for the tested cosmological model. Building on the Fisher forecast that I presented in section 3.8, the BNT transform should be used to perform the naturally following work which would then include realistic source distributions and a tomographic analysis. This would allow us to precisely evaluate from which scales the information is coming, and thus eventually explore how careful one would need to be with cosmological constraints in lensing if the vast majority of the information was found to come from not very well known non-linear scales. Continuing on the subject of a tomographic analysis with the help of the BNT transform, the second part of chapter 4 strongly advocates for the switch from the PDF as an observable to the CGF. Indeed, though their information content and observational systematics are strictly the same, I have demonstrated that it is much more easily computed in the case of a joint analysis between nulled redshift bins. However, in practice, some questions regarding a practical implementation remain open and need to be tackled. Those essentially concern how to translate cuts in the tails of the PDFs and binning of the $\kappa/M_{\rm ap}$ values in terms of cuts and binning in λ which is not trivial and necessitates some exploration and experimentation.

As for projects that re-think a bit what we have been doing, there is some work to be carried out on the extension of the large deviation principle outside of the large deviations regime for the matter density field itself. I have three points in mind, first replacing the spherical collapse by a more general density-dependent collapse (for example ellipsoidal) which would reduce to the spherical collapse close to the mean density. Then working on the differentiability assumption for the SCGF, does it translate into the prediction of a novel signature in the tails of the PDF or would it just indicate the limit of the extension of the spherical collapse dynamics outside the regime of large deviations? Finally, the extension of the PDF of the random vector made of the density field filtered in top-hats of different radii outside the large deviations regime is not extremely satisfying, mainly because this forces us to choose only one driving parameter while we would like to fix the variance of each density and all their covariances. This would require to re-write the large deviation principle for this random vector in terms of different "speed" (roughly several driving parameters), which is technically feasible but requires more work to precisely understand how to do it.

Finally, all advancements in the field of numerical simulations for weak-lensing will benefit our future works as they still allow us to delimit the validity regime of our theoretical developments. I hope in return that our own advancements can also help to test those simulations, the two methods being intertwined and co-dependent to improve our knowledge on the physics of the gravitational instability. Indeed, I hope that this work actually high-

lights the importance of developing theoretical models even when one could think of ways to simply measure things in numerical simulations. They are not immune to artefacts and a general caution towards these tools, though essential in our fields, is I think appropriate. Only then, with a balanced mix between the theoretical understanding and the numerical modelling, will we be able to rise to the challenge of elucidating the mysteries surrounding, all other things being equal, dark energy through the study of the large scale structure of our Universe.

Some of the mentioned future projects are already in preparation, some are still speculative, nevertheless I would like to convey, as a final word in this PhD dissertation what an active and interesting field of research this is, and what an honour it has been to discover it and work on it with such brilliant researchers.



A Appendices

Contents

A.1. Mathematical tools and definitions

A.1.1 Fourier conventions

Two Fourier conventions are used throughout this thesis. In the first chapter, because it allows for a definition of the power spectrum without additional πs in already large expressions, I choose

$$f(x) = \int \frac{\mathrm{d}k}{(2\pi)^{1/2}} \tilde{f}(k) e^{ik\cdot x}. \tag{A.1}$$

In the following chapters, I switch back to the more usual (at least in astrophysics) convention because I make use of the CLASS code, which outputs power spectra computed with

$$f(x) = \int \frac{\mathrm{d}k}{2\pi} \tilde{f}(k) e^{ik\cdot x}. \tag{A.2}$$

A.1.2 Generating functions and PDF

Throughout this work, I make use of different statistical quantities that I briefly introduce here for clarity. From the PDF $\mathcal{P}_X$ of some continuous random variable X one can define the moment generating function as the Laplace transform of the PDF

$$M_X(\lambda) = \mathrm{E}\left(e^{\lambda X}\right) = \int_{-\infty}^{+\infty} e^{\lambda x}\mathcal{P}_X(x)\mathrm{d}x, \tag{A.3}$$

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

$$\begin{aligned} M_X(\lambda) = \mathrm{E}\left(e^{\lambda X}\right) &= 1 + \lambda \mathrm{E}(X) + \frac{\lambda^2 \mathrm{E}\left(X^2\right)}{2!} + \frac{\lambda^3 \mathrm{E}\left(X^3\right)}{3!} + \cdots \\ &= \sum_{n=0}^{+\infty} \frac{\lambda^n \mathrm{E}\left(X^n\right)}{n!}, \end{aligned} \tag{A.4}$$

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

$$\phi_X(\lambda) = \log(M_X(\lambda)) = \sum_{n=1}^{+\infty} k_n \frac{\lambda^n}{n!} \tag{A.5}$$

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

These definitions can of course be extended to the case of joint probabilities. For example, for the case of two continuous random variables X and Y, the joint cumulant generating function reads

$$\phi_{X,Y}(\lambda_1, \lambda_2) = \log(M_{X,Y}(\lambda_1, \lambda_2)) = \log\left[\mathrm{E}\left(e^{\lambda_1 X + \lambda_2 Y}\right)\right] \tag{A.6}$$

$$= \sum_{p,q=0}^{+\infty} \langle X^p Y^q \rangle_c \frac{\lambda_1^p \lambda_2^q}{p!q!} - 1, \tag{A.7}$$

which in particular allows us to straightforwardly define the CGF of any linear combination of random variables from their joint CGF which is useful in our context. More precisely, noticing that $\phi_{X+Y}(\lambda) = \phi_{X,Y}(\lambda, \lambda)$ allows us to generalise the famous relation $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$ to any cumulant

$$k_{n,X+\alpha Y} = \sum_{j=0}^{n} \binom{n}{j} \alpha^{n-j} \langle X^j Y^{n-j} \rangle_c. \tag{A.8}$$

*Note that I make use throughout this work, when making measurements, of the ergodicity hypothesis where one assumes that ensemble averages are equivalent to spatial averages ($E(.) \to \langle . \rangle$) over one realisation of a random field at one fixed time. This requires that spatial correlations decay sufficiently rapidly with separation such that one has access to many statistically independent volumes in one realisation.

Since we are over the course of this thesis interested in the aperture mass and thus the density slope between two concentric disks along a line of sight, I also note that the CGF of the difference of two random variables is simply expressed from their joint CGF through

$$\phi_{\delta_2-\delta_1}(\lambda) = \phi_{\delta_1,\delta_2}(-\lambda, \lambda). \tag{A.9}$$

It turns out that the quantities

$$S_n = \frac{k_n}{k_2^{n-1}}, \tag{A.10}$$

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

$$\varphi_X(\lambda) = \lim_{k_2\to 0} \sum_{n=0}^{+\infty} S_n \frac{\lambda^n}{n!} = \lim_{k_2\to 0} k_2\, \phi_X\left(\frac{\lambda}{k_2}\right), \tag{A.11}$$

that I in our context extrapolate to non-zero values of the variance. Eventually, one can then reconstruct the PDF for the random variable X as an inverse Laplace transform (inverting equation (A.3)) given by

$$\mathcal{P}_X(x) = \int_{-i\infty}^{+i\infty} \frac{\mathrm{d}\lambda}{2\pi i} \exp\left(-\lambda x + \phi_X(\lambda)\right). \tag{A.12}$$

A.1.3 Gram–Charlier A expansion

The Gram–Charlier A series (and for that matter the Edgeworth series) are series that approximate a probability distribution in terms of its cumulants. Formally the series are identical and the names are used rather indistinguishably in the astrophysics literature – I present what is usually meant – but the arrangement of the terms and thus the accuracy of truncating the series differ. The idea behind these expansions is actually quite similar to the type of computations we have performed in the context of the large deviation principle for the density field – we start from a restriction of the moment generating function and recover the PDF through an inverse transform.

The *characteristic function* of a random variable X is the moment generating function restricted to the imaginary axis* and thus the Fourier transform of the PDF. Thus, still denoting the cumulants by k_n it is expressed as

$$\hat{M}_X(t) = \exp\left[\sum_{n=1}^{\infty} k_n \frac{(it)^n}{n!}\right]. \tag{A.13}$$

If I now introduce another random variable Y with cumulants γ_n then the formal identity

$$\hat{M}_X(t) = \exp\left[\sum_{n=1}^{\infty} (k_n - \gamma_n) \frac{(it)^n}{n!}\right] \hat{M}_Y(t) \tag{A.14}$$

*With the addition that the characteristic function exists even when the PDF and the moments/cumulants do not.

is valid. Then noticing that $(it)^n$ is the differential operator $(-1)^n\partial^n$ in real space we have

$$\mathcal{P}_X(x) = \exp\left[\sum_{n=1}^{\infty}(k_n - \gamma_n)\frac{(-\partial)^n}{n!}\right]\mathcal{P}_Y(x). \tag{A.15}$$

Now taking Y to be Gaussian distributed, with the same mean $k_1 = \gamma_1 = \mu$ and variance $k_2 = \gamma_2 = \sigma^2$ as X I finally arrive at the result

$$\mathcal{P}_X(x) = \exp\left[\sum_{n=3}^{\infty}k_n\frac{(-\partial)^n}{n!}\right]\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{A.16}$$

The rest is cosmetic. By definition of the Bell polynomials and the Hermite polynomials we have

$$\exp\left[\sum_{n=3}^{\infty}k_n\frac{(-\partial)^n}{n!}\right] = \sum_{n=0}^{\infty}B_n(0,0,k_3,\ldots,k_n)\frac{(-\partial)^n}{n!} \tag{A.17}$$

and

$$\mathcal{P}_Y^{(n)}(x) = \frac{(-1)^n}{\sigma^n}H_{en}\left(\frac{x-\mu}{\sigma}\right)\mathcal{P}_Y(x). \tag{A.18}$$

This gives me the final expression for the Gram–Charlier A expansion, which is interpreted in terms of corrections to the Gaussian distribution from other cumulants as

$$\mathcal{P}_X(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\sum_{n=0}^{\infty}\frac{1}{n!\sigma^n}B_n(0,0,k_3,\ldots,k_n)H_{en}\left(\frac{x-\mu}{\sigma}\right). \tag{A.19}$$

Note that any truncation of this expression is not at all guaranteed to be positive and as such is not a proper PDF. Note that the entire series may also diverge in many cases as $\mathcal{P}_X(x)$ has to fall faster than $\exp(-x^2/4)$ at infinity.

Looking at the first correction we get

$$\mathcal{P}_X(x) \simeq \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\left[1 + \frac{k_3}{3!\sigma^3}H_{e3}\left(\frac{x-\mu}{\sigma}\right)\right] \tag{A.20}$$

with $H_{e3}(x) = x^3 - 3x$.

For completeness, let me finally mention that the Edgeworth expansion consists of the same result but this time seen as an extension of the central limit theorem, or rather a way to estimate the error made when assuming that a distribution is Gaussian under said theorem. In this sense it makes the link with large deviation theory especially clear.

Considering now that X is the standardised sum of n identical continuous random variables Z_i with mean μ, variance σ^2 and higher cumulants k_r

$$X_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Z_i - \mu}{\sigma}, \tag{A.21}$$

and the cumulants of X_n are given by additivity by

$$k_r^{X_n} = \frac{nk_r}{n^{r/2}\sigma^r} = \frac{\lambda_r}{n^{r/2-1}}. \tag{A.22}$$

Then the Gram-Charlier A expansion of X_n reads

$$\mathcal{P}_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sum_{r=0}^{\infty} \frac{1}{r!} B_r(0, 0, \frac{\lambda_3}{n^{1/2}}, \ldots, \frac{\lambda_r}{n^{r/2-1}}) H_{er}(x), \tag{A.23}$$

and now terms are collected according to powers of n. This changes the order of the terms from the usual Gram-Charlier A expansion but does not change the fact that the first term, that comes from the skewness, is the dominant correction. In our context, given that our driving parameter is $n = 1/\sigma^2$ we could have already obtained the Edgeworth expansion from the Gram-Charlier A series also expanding in σ and collecting terms this way.

The generalisation of the Gram-Charlier A expansion to multivariate series is actually straightforward in the sense that almost the same steps are followed. However, the notation is a bit more involved and its derivation is thus not particularly useful in the context of this thesis. A derivation can be found in Bhaveshkumar C (2015) and the results are stated in their equations (42) and (43).

A.2. Deformation matrix from the fermat principle

Let me start by noting that at the level of a flat FLRW background, the infinitesimal comoving separation distance in some direction $\boldsymbol{\theta}$ at some comoving radial distance χ $\boldsymbol{x}(\boldsymbol{\theta}, \chi)$ between two geodesics converging at the observer is equal to the comoving radial distance χ as stated in equation (1.25). This implies that

$$\frac{\mathrm{d}^2\boldsymbol{x}}{\mathrm{d}\chi^2} = 0, \tag{A.24}$$

which could also be seen as (1.10) written for the background. I now want to include some density perturbations in the propagation equation (A.24). If one assumes that the Newtonian potential Φ is small compared to c^2 and that the typical scales over which Φ changes significantly are much smaller than the curvature scale of the background then according to Fermat's principle (which directly comes from the extremisation of the action of light moving through curved space-time) each light ray is deflected from its unperturbed path by the transverse gradient of the potential it passes through. Since any physical fiducial ray will be deflected by potential gradients along its way the right-hand side of equation (A.24) must thus contain the difference of the perpendicular potential gradients between the two rays to account for the relative deflection of the two rays

$$\frac{\mathrm{d}^2\boldsymbol{x}}{\mathrm{d}\chi^2} = -\frac{2}{c^2}\Delta[\nabla_\perp \Phi(\boldsymbol{x}(\boldsymbol{\theta}, \chi), \chi)]. \tag{A.25}$$

This can be solved in a similar fashion as the optical Sachs equation (1.2) in the main text and one gets

$$\boldsymbol{x}(\boldsymbol{\theta}, \chi_s) = \mathcal{D}_0(z_s)\boldsymbol{\theta} - \frac{2}{c^2}\int_0^{\chi_s} \mathrm{d}\chi \mathcal{D}_0(z_s - z)\Delta[\nabla_\perp \Phi(\boldsymbol{x}(\boldsymbol{\theta}, \chi), \chi)]. \tag{A.26}$$

Note that contrary to the main text, the transverse gradient operator is here expressed in comoving coordinates. Then going back to the deformation matrix $A_{ij} = \mathcal{D}_{ij}/\mathcal{D}_0$ one gets

$$A_{ij}(\boldsymbol{\theta}, \chi_s) = \delta^D_{ij} - \frac{2}{c^2}\int_0^{\chi_s} \mathrm{d}\chi \frac{\mathcal{D}_0(z_s - z)\mathcal{D}_0(z)}{\mathcal{D}_0(z_s)} \Phi_{,ik}(\boldsymbol{x}(\boldsymbol{\theta}, \chi), \chi) A_{kj}(\boldsymbol{\theta}, \chi). \tag{A.27}$$

The last expression is then explicitly obtained by expanding in parallel $A_{ij} = \delta^D_{ij} + A^{(1)}_{ij} + ...$ and $\boldsymbol{x}(\boldsymbol{\theta}, \chi_s)$ in powers of Φ. This last expansion corresponds to dropping the Born-approximation since the zeroth order of $\boldsymbol{x}(\boldsymbol{\theta}, \chi_s)$ gives in the previous integral the unperturbed light trajectory while further expanding accounts for the "real" position of lenses. The details of the computation are very similar to what was done in the main text. After some algebra, one eventually finds

$$A^{(2)}_{ij}(\boldsymbol{\theta}, \chi_s) = \frac{4}{c^4}\int_0^{\chi_s} \mathrm{d}\chi \frac{\mathcal{D}_0(z_s - z)\mathcal{D}_0(z)}{\mathcal{D}_0(z_s)} \int_0^{\chi} \mathrm{d}\chi' \mathcal{D}_0(z - z') \left[\Phi_{,ijl}(\mathcal{D}_0(\chi))\Phi_{,l}(\mathcal{D}_0(\chi')) + \frac{\mathcal{D}_0(\chi')}{\mathcal{D}_0(\chi)}\Phi_{,ik}(\chi)\Phi_{,kj}(\chi')\right], \tag{A.28}$$

which is strictly equivalent to what was found in the main text.

A.3. Spherical collapse dynamics

A.3.1 Einstein-de-Sitter universe

I here consider a spherically symmetric over-density perturbation in a spatially flat universe filled with non-relativistic matter ($\Omega_m = 1$). I then study the motion of a spherical shell of matter of radius y enclosing the total mass M. Before any shells cross each other, the mass within the shell is conserved and due to Birkhoff's theorem the mechanical energy $\varepsilon < 0$ – negative because the system is gravitationally bound and thus collapsing – of the shell is conserved so that the kinetics energy and the gravitational potential energy compensate each other

$$\frac{1}{2}\left(\frac{\mathrm{d}y}{\mathrm{d}\tau}\right)^2 - \frac{GM}{y} = \varepsilon \tag{A.29}$$

where τ is the physical time. The parametric solution of this differential equation is obtained as

$$\begin{cases} y = -\dfrac{GM}{2\varepsilon}(1 - \cos\theta) \\ \tau = \dfrac{GM}{(-2\varepsilon)^{3/2}}(\theta - \sin\theta). \end{cases} \tag{A.30}$$

I then want to move to the comoving radius of the shell $R = y/a(\tau)$ and to the scale factor $a(\tau)$. In an EdS universe we have

$$a = \left(\frac{8\pi G}{3}\rho_i a_i^3\right)^{1/3}\left(\frac{3}{2}\tau\right)^{2/3}, \tag{A.31}$$

and the mass enclosed within the shell can be written as

$$M = \frac{4\pi}{3}\rho_i a_i^3 r^3, \tag{A.32}$$

where ρ_i, a_i and r are the matter density, the scale factor and the comoving radius of the shell at some early time when the universe was almost homogeneous and r has a finite limit at $a_i \to 0$ which coincides with the Lagrangian radius of the shell. We thus obtain

$$\begin{cases} R = r\left(\frac{2}{9}\right)^{1/3} \frac{1-\cos\theta}{(\theta-\sin\theta)^{2/3}} \\ a = \left(\frac{9}{2}\right)^{1/3} \frac{4\pi G}{3(-2\varepsilon)}\rho_i a_i^3 r^2 (\theta-\sin\theta)^{2/3}. \end{cases} \tag{A.33}$$

From the definition of the matter density contrast enclosed in the shell we have

$$1+\delta(R) = \frac{3M}{4\pi y^3 \rho_{\rm univ}} = \left(\frac{r}{R}\right)^3 \tag{A.34}$$

where $\rho_{\rm univ} = \rho_i a_i / a^3$ is the mean density in the universe and this gives us the relation between the Eulerian and Lagragian radius of the shell which illustrates mass conservation. From (A.33), I thus obtain the density contrast within the shell as a function of the development angle θ

$$\delta = \frac{9}{2}\frac{(\theta-\sin\theta)^2}{(1-\cos\theta)^3} - 1. \tag{A.35}$$

Now all that remains is to make the link with the evolution of the linear density contrast within the shell. For that we need to relate the mechanical energy ε to the initial over-density. When θ is small we can Taylor expand the trigonometric functions and arrive at

$$\begin{cases} a_i = \frac{\theta_i^2}{2}\frac{4\pi G}{3(-2\varepsilon)}\rho_i a_i^3 r^2 \\ \delta_i(r) = \frac{3}{20}\theta_i^2 \end{cases} \tag{A.36}$$

which finally gives

$$\varepsilon = -\frac{5}{3}\frac{\delta_i(r)}{a_i}\frac{4\pi G}{3}\rho_i a_i^3 r^2. \tag{A.37}$$

Thus recalling that the linear density contrast in an EdS universe evolves as $\delta_l = a\delta_i a_i$ I finally obtain

$$\delta_l = \frac{3}{20}[6(\theta-\sin\theta)]^{2/3}. \tag{A.38}$$

Equations (A.35) and (A.38) thus define the mapping between the linear and non-linear over-density contrast in the spherical collapse dynamics in an EdS universe. For the case of an under-density, the same steps are performed and we arrive at

$$\begin{cases} \delta = \frac{9(\mathrm{sh}\,\theta-\theta)^2}{2(\mathrm{ch}\,\theta-1)^3} - 1 \\ \delta_l = -\frac{3}{20}[6(\mathrm{sh}\,\theta-\theta)]^{2/3}. \end{cases} \tag{A.39}$$

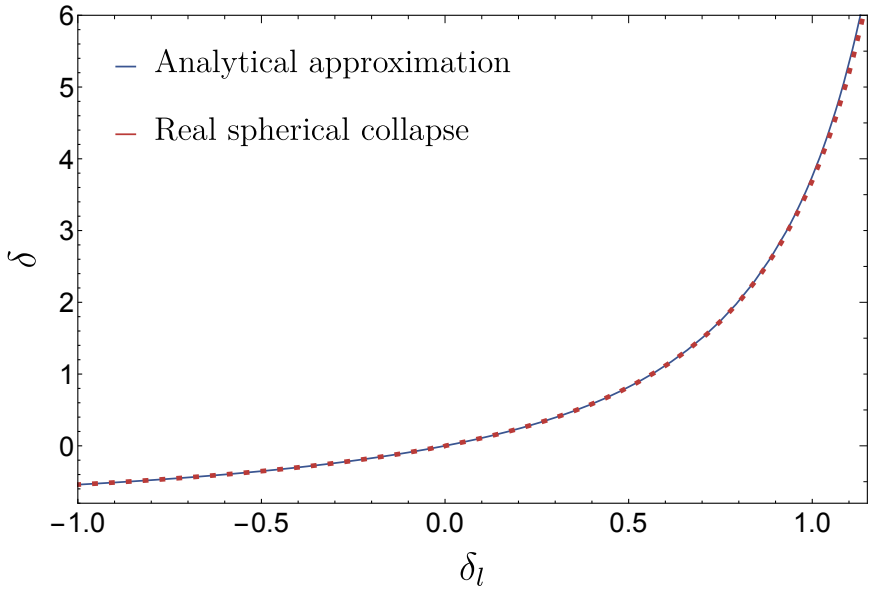


Figure A.1: Comparison of the analytical parametrisation and the real spherical collapse mapping between the linear and non-linear matter density contrast within the collapsing sphere before shell-crossing.

To conclude this subsection, let me now illustrate in figure A.1 the accuracy of the parametrisation I used throughout this thesis. The parametrisation is given by

$$\delta = \left(1 - \frac{\delta_l}{\nu}\right)^{-\nu} - 1, \tag{A.40}$$

with $\nu = 21/13$ in 3D chosen so as to reproduce the quadratic term of the Taylor expansion of δ for small δ_l which has an impact on the skewness. Globally the analytical approximation is very accurate except maybe for very high values of the non-linear density contrast for which our calculations might not be accurate anyway since the spherical collapse model as a proxy for the non-linear gravitational instability might not be accurate anymore.

A.3.2 Link with standard perturbation theory

Long before our works involving the spherical collapse as the most probable dynamics driving the evolution of the density contained within a sphere in the context of the large deviation principle, Bernardeau (1992) had already illustrated how spherical collapse dynamics actually drives most* of the formation of structures at leading (tree)-order in Eulerian perturbation theory. I do no want here to just re-write the details of the calculations which are clear enough in the original paper but let me just remind the idea.

The first step is to re-write equation (A.29) changing variables from the physical radius y to the density within the shell $\rho = (r/R)^N$ and from the physical time to the linear density field $\delta_l \equiv a(\tau)\delta_i/a_i$. This is done easily still in an EdS setting and one finds that the equation is written as

$$-\rho\delta_l^2\rho'' + \frac{N+1}{N}\left(\delta_l\rho'\right)^2 - \frac{3}{2}\rho\delta_l\rho' + \frac{3}{2}\rho^2\left(\rho - 1\right) = 0 \tag{A.41}$$

*The spherical collapse gives the exact reduced cumulants and not the cumulants themselves

and whose solution is given by equations (A.35) and (A.38) for an initial over-density and (A.39) for under-densities, the prime denotes a derivation by δ_l and N is the dimension of the collapse ($N = 1$ planar, $N = 2$ cylindrical, $N = 3$ spherical etc). Note that though I personally find clearer the demonstration of the spherical collapse dynamics starting from equation (A.29), obtaining the previous equation is however more natural starting from the one-fluid Newtonian system of equations, namely the Euler, continuity and Poisson equations and reducing it to the case of spherical symmetry for the density field assuming no rotation and no shear as is done in section 6.4 of Mukhanov (2005).

The idea is then to demonstrate that the generating functions of the perturbation theory vertices follow this exact same equation so that its solution is also known. I thus now want to reiterate what these vertices are and how they are related to the statistics of the density field. Starting from the Fourier expression of the one-fluid Newtonian system of equations and assuming that the density field (as well as the divergence of the gravitational potential) can be expanded order by order in a series

$$\delta(\boldsymbol{x}, t) = \sum_n \delta^{(n)}(\boldsymbol{x}, t) \tag{A.42}$$

where $\delta^{(1)}$ is the linear density field, one can show that each order of the density field is obtained as

$$\delta^{(n)}(\boldsymbol{x}, t) = D(t)^n \int \frac{\mathrm{d}\boldsymbol{k}_1}{(2\pi)^N} e^{i\boldsymbol{k}_1\cdot\boldsymbol{x}} \delta^{(1)}(\boldsymbol{k}_1) \dots \frac{\mathrm{d}\boldsymbol{k}_n}{(2\pi)^N} e^{i\boldsymbol{k}_n\cdot\boldsymbol{x}} \delta^{(1)}(\boldsymbol{k}_n) F_n(\boldsymbol{k}_1 \dots \boldsymbol{k}_n). \tag{A.43}$$

$D(t)$ is the linear growth factor (in cosmologies where the time-dependence of the linear field can be split from the scale dependence) and F_n are the so called perturbation theory kernels, which are defined with some recursion relations found in many references, see for example Bernardeau et al. (2002). As a consequence, every unsmoothed moment of the density field $\langle\delta^p\rangle$ at leading order only features kernels of degree at most $p-1$. For Gaussian initial conditions and thanks to Wick's theorem, every moment of the density field is thus expressed as an integral of the F_n kernels and the linear power spectrum which only depends on the radial components of the wave vectors. We thus define the perturbation theory vertices as the angular integration of the perturbation theory kernels as

$$\nu_n = n! \int \mathrm{d}\Omega_1 \dots \mathrm{d}\Omega_n F_n(\boldsymbol{k}_1 \dots \boldsymbol{k}_n), \tag{A.44}$$

and interestingly, the leading order unsmoothed reduced cumulants of the density field $S_p = \langle\delta^p\rangle_c / \sigma_l^{2(p-1)}$ are then given as some combination of $\nu_{i\leq p-1}$. Thus the knowledge of every ν_n gives us every cumulant of the density field at leading order in perturbation theory. The advance of Bernardeau (1992) is thus to re-write the one-fluid Newtonian system of equations so as to show that the generating function of the ν_n – that I here write on purpose $\rho(\delta_l) = \sum_n \nu_n \delta_l^n / n!$ with $\rho = 1 + \delta_l$ for small δ_l satisfies equation (A.41). As such the spherical collapse dynamics is the leading order dynamics in Eulerian perturbation theory.

A.3.3 ΛCDM universe

In the presence of a cosmological constant, equation (A.29) becomes

$$\frac{1}{2}\left(\frac{dy}{d\tau}\right)^2 - \frac{GM}{y} - \frac{\Lambda y^2}{6} = \varepsilon \tag{A.45}$$

which unfortunately does not admit analytical solutions and has to be integrated numerically. To easily obtain the mapping between the linear and the non-linear density contrast, it is easier to switch to variables $\xi = r/R$ and the scale factor a. One thus uses

$$\frac{1}{a^2}\left(\frac{da}{d\tau}\right)^2 = \frac{8\pi G}{3}\rho_{\rm univ} + \frac{\Lambda}{3} \tag{A.46}$$

with $\rho_{\rm univ} = \rho_0/a^3$, $\Lambda = 8\pi G\rho_0\Omega_\Lambda/\Omega_m$ and ρ_0 is the average matter density in the universe today. The mechanical energy is expressed noticing that the mass conservation still applies and $1+\delta = \xi^3$ and matching this expression at early times to linear terms in a in the differential equation when $a \to 0$. Finally we express the initial overdensity as the linear density field noticing that $\delta_i/a_i = C/D(a)\delta_l$ where C is a constant that comes from the fact that the linear growth factor is normalised so that $D(0) = 1$ which leads to an offset between D and a at high redshift in the matter dominated era (EdS universe). I thus arrive at

$$\left(1 - \frac{d\ln\xi}{d\ln a}\right)^2 = \left(1 + \frac{\Omega_\Lambda}{\Omega_m}a^3\right)^{-1}\left(\xi^3 - \frac{5C}{3D(a)}\delta_l a\xi^2 + \frac{\Omega_\Lambda}{\Omega_m}a^3\right). \tag{A.47}$$

To solve this equation, one fixes a value of δ_l at $a = 1/(1+z)$, solves the equation from $a = 0$ to a with initial condition $\xi_{a=0} = 1$, and finally computes $\delta = \xi^3(a) - 1$.

A.4. Likelihood function for moments of the convergence field

In figure 3.19, I show that a large part of the information contained in the PDF comes from the second and third moments alone. This is to be treated with some caution, however, as we remove information from the PDF by truncating the tails, while the full distribution of values is included in our measurements of the moments. The moments-only constraints also benefit from a smaller Kaufman-Hartlap factor (see equation 3.66). In addition, constraints with a finite number of moments can only perform well when the total number of observables (in the case of figure 3.19, two scales and two moments, so four) is greater than the number of parameters being constrained.

Furthermore, we can highlight a particular advantage of performing Fisher matrix calculations with the bulk of the PDF instead of individual moments: the individual moments show very non-Gaussian distributions (numerical Likelihood functions) when measured from simulations with small patch sizes. Figure A.2 shows the distributions of the third and fourth moments measured at two smoothing scales measured from patches of FLASK-generated maps with side length 5 deg. This is not the case for the bulk of the PDF as seen in figure 3.15.

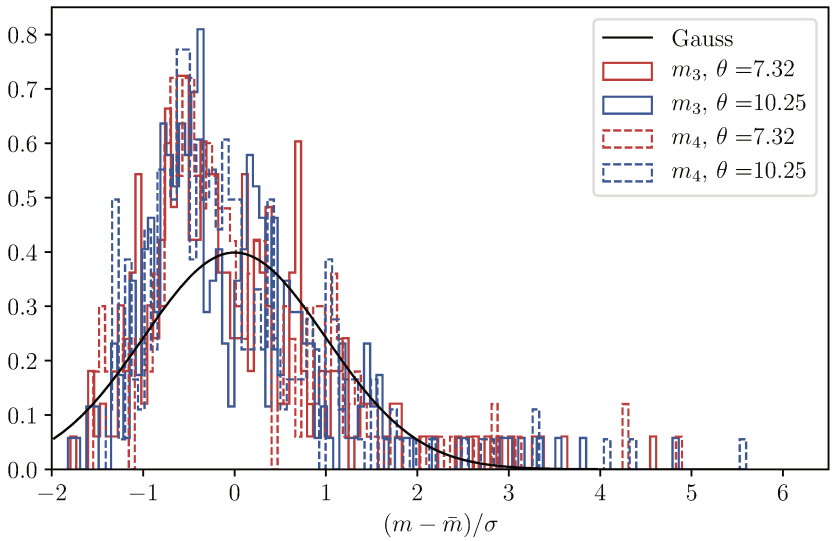


Figure A.2: The distributions of the third and fourth central moments taken from 256 25 deg^2 patches from FLASK-generated maps. This figure was published in Boyle et al. (2021).

A.5. Rewriting the trivariate PDF as functional of bivariate PDFs

I aim here at expressing $\mathcal{P}(\hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_3)$ purely as a function of $\mathcal{P}(\hat{\kappa}_1, \hat{\kappa}_2)$ and $\mathcal{P}(\hat{\kappa}_2, \hat{\kappa}_3)$. We know this is technically possible given equation (4.35).

I start by writing the CGFs used as building blocks for the joint multivariate PDF in terms of univariate and bivariate PDFs using Laplace transformations as

$$e^{\hat{\phi}(\lambda_2)} = \int d\hat{\kappa}_2' \exp(\hat{\kappa}_2'\lambda_2)\mathcal{P}(\hat{\kappa}_2')\,, \tag{A.48}$$

$$e^{\hat{\phi}_{\kappa_1}(\lambda_2)} = \int d\hat{\kappa}_2' \exp(\hat{\kappa}_2'\lambda_2)\mathcal{P}(\hat{\kappa}_1, \hat{\kappa}_2')\,, \tag{A.49}$$

and similarly for $\exp(\hat{\phi}_{\kappa_3}(\lambda_2))$. I insert these into the expression for the trivariate PDF built from pairwise CGFs from equation (4.36) and thus finally obtain

$$\mathcal{P}(\hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_3) = \int \frac{d\lambda_2}{2\pi\mathrm{i}} e^{-\hat{\kappa}_2\lambda_2} \frac{\int d\hat{\kappa}_2' e^{\hat{\kappa}_2'\lambda_2}[\mathcal{P}(\hat{\kappa}_1, \cdot) * \mathcal{P}(\cdot, \hat{\kappa}_3)](\hat{\kappa}_2')}{\int d\hat{\kappa}_2'' e^{\hat{\kappa}_2''\lambda_2}\mathcal{P}(\hat{\kappa}_2'')} \tag{A.50}$$

where $\mathcal{P}(\hat{\kappa}_1, \cdot) * \mathcal{P}(\cdot, \hat{\kappa}_3)$ denotes the convolution of $\mathcal{P}(\hat{\kappa}_1, \hat{\kappa}_2)$ and $\mathcal{P}(\hat{\kappa}_2, \hat{\kappa}_3)$ w.r.t. $\hat{\kappa}_2$. This is a result of the convolution theorem stating that the product of the Laplace transforms of two functions can be written as the Laplace transform of their convolution. As expected from the somewhat complicated relationship between PDFs and CGFs, equation (A.50) is unfortunately not nearly as simple as equation (4.35) though they both depict the exact same result without any loss of information.

Bibliography

Alam S., et al., 2017, The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample, MNRAS, 470, 2617

Amendola L., et al., 2018, Cosmology and fundamental physics with the Euclid satellite, Living Reviews in Relativity, 21

Baker T., Clampitt J., Jain B., Trodden M., 2018, Void lensing as a test of gravity, Phys. Rev. D, 98, 023511

Balian R., Schaeffer R., 1989, Scale-invariant matter distribution in the universe., Astronomy and Astrophysics, 220, 1

Bartelmann M., Schneider P., 2001, Weak gravitational lensing, Phys. Rep., 340, 291

Barthelemy A., Codis S., Uhlemann C., Bernardeau F., Gavazzi R., 2020a, A nulling strategy for modelling lensing convergence in cones with large deviation theory, MNRAS, 492, 3420

Barthelemy A., Codis S., Bernardeau F., 2020b, Post-Born corrections to the one-point statistics of (CMB) lensing convergence obtained via large deviation theory, MNRAS, 494, 3368

Barthelemy A., Bernardeau F., Codis S., Uhlemann C., 2021a, Numerical complexity of the joint nulled weak-lensing probability distribution function, arXiv e-prints, p. arXiv:2106.11632

Barthelemy A., Codis S., Bernardeau F., 2021b, Probability distribution function of the aperture mass field with large deviation theory, MNRAS, 503, 5204

Battye R. A., Charnock T., Moss A., 2015, Tension between the power spectrum of density perturbations measured on large and small scales, Physical Review D, 91, 103508

Baugh C. M., Gaztanaga E., Efstathiou G., 1995, A comparison of the evolution of density fields in perturbation theory and numerical simulations - II. Counts-in-cells analysis, MNRAS, 274, 1049

Beck D., Fabbian G., Errard J., 2018, Lensing reconstruction in post-Born cosmic microwave background weak lensing, Phys. Rev. D, 98, 043512

Bel J., et al., 2016, The VIMOS Public Extragalactic Redshift Survey (VIPERS). On the recovery of the count-in-cell probability distribution function, Astronomy and Astrophysics, 588, A51

Bernardeau F., 1992, The Gravity-induced Quasi-Gaussian Correlation Hierarchy, ApJ, 392, 1

Bernardeau F., 1994, The Nonlinear Evolution of Rare Events, The Atrophysical Journal, 427, 51

Bernardeau F., 1995, The angular correlation hierarchy in the quasilinear regime., Astronomy and Astrophysics, 301, 309

Bernardeau F., Kofman L., 1995, Properties of the Cosmological Density Distribution Function, ApJ, 443, 479

Bernardeau F., Reimberg P., 2016, Large deviation principle at play in large scale structure cosmology, Phys. Rev. D, 94, 063520

Bernardeau F., Valageas P., 2000, Construction of the one-point PDF of the local aperture mass in weak lensing maps, Astronomy and Astrophysics, 364, 1

Bernardeau F., van Waerbeke L., Mellier Y., 1997, Weak lensing statistics as a probe of Ω and power spectrum., Astronomy and Astrophysics, 322, 1

Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Large-scale structure of the Universe and cosmological perturbation theory, Phys. Rep., 367, 1

Bernardeau F., Bonvin C., Vernizzi F., 2010, Full-sky lensing shear at second order, Phys. Rev. D, 81, 083002

Bernardeau F., Pichon C., Codis S., 2014a, Statistics of cosmic density profiles from perturbation theory, Phys. Rev. D, 90, 103519

Bernardeau F., Nishimichi T., Taruya A., 2014b, Cosmic shear full nulling: sorting out dynamics, geometry and systematics, MNRAS, 445, 1526

Betoule M., et al., 2014, Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples, Astronomy and Astrophysics, 568, A22

Bhaveshkumar C D., 2015, Multivariate Generalized Gram-Charlier Series in Vector Notations, arXiv e-prints, p. arXiv:1503.03212

Blas D., Lesgourgues J., Tram T., 2011, The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes, Journal of Cosmology and Astroparticle Physics, 2011, 034–034

Blazek J. A., MacCrann N., Troxel M. A., Fang X., 2019, Beyond linear galaxy alignments, Physical Review D, 100, 103506

Böhm V., Schmittfull M., Sherwin B. D., 2016, Bias to CMB lensing measurements from the bispectrum of large-scale structure, Phys. Rev. D, 94, 043519

Böhm V., Sherwin B. D., Liu J., Hill J. C., Schmittfull M., Namikawa T., 2018, Effect of non-Gaussian lensing deflections on CMB lensing measurements, Phys. Rev. D, 98, 123510

Böhm V., Modi C., Castorina E., 2019, Lensing corrections on galaxy-lensing cross correlations and galaxy-galaxy auto correlations, arXiv e-prints, p. arXiv:1910.06722

Boyle A., Uhlemann C., Friedrich O., **Barthelemy A.**, Codis S., Bernardeau F., Giocoli C., Baldi M., 2021, Nuw CDM cosmology from the weak lensing convergence PDF, MNRAS,

Breton M.-A., Rasera Y., Taruya A., Lacombe O., Saga S., 2019, Imprints of relativistic effects on the asymmetry of the halo cross-correlation function: from linear to non-linear scales, MNRAS, 483, 2671

Carrasco J. J. M., Hertzberg M. P., Senatore L., 2012, The effective field theory of cosmological large scale structures, Journal of High Energy Physics, 2012, 82

Cataneo M., Uhlemann C., Arnold C., Gough A., Li B., Heymans C., 2021, The matter density PDF for modified gravity and dark energy with Large Deviations Theory, arXiv e-prints, p. arXiv:2109.02636

Clerkin L., et al., 2017, Testing the lognormality of the galaxy and weak lensing convergence distributions from Dark Energy Survey maps, MNRAS, 466, 1444

Codis S., et al., 2015, Intrinsic alignment of simulated galaxies in the cosmic web: implications for weak lensing surveys, MNRAS, 448, 3391

Codis S., Bernardeau F., Pichon C., 2016, The large-scale correlations of multicell densities and profiles: implications for cosmic variance estimates, MNRAS, 460, 1598

Coe D., et al., 2019, RELICS: Reionization Lensing Cluster Survey, Astrophysical Journal, 884, 85

Coles P., Jones B., 1991, A lognormal model for the cosmological mass distribution, MNRAS, 248, 1

Colombi S., 1994, A "skewed" lognormal approximation to the probability distribution function of the large-scale density field., ApJ, 435, 536

Crittenden R. G., Turok N., 1996, Looking for a Cosmological Constant with the Rees-Sciama Effect, Physical Review Letters, 76, 575

DES Collaboration et al., 2021, Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing, arXiv e-prints, p. arXiv:2105.13549

DES Collaboration 2018, Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing, Physical Review D, 98

Deshpande A. C., et al., 2020, Euclid: The reduced shear approximation and magnification bias for Stage IV cosmic shear experiments, Astronomy and Astrophysics, 636, A95

Desjacques V., Jeong D., Schmidt F., 2018, Large-scale galaxy bias, Physics Reports, 733, 1

Dodelson S., Schneider M. D., 2013, The effect of covariance estimator error on cosmological parameter constraints, Physical Review D, 88, 063537

Dodelson, S. and Schmidt, F. 2020, Modern Cosmology - 2nd Edition. Elsevier

Ellis R., 1985, Entropy, Large Deviations, and Statistical Mechanics. Grundlehren der mathematischen Wissenschaften, Springer-Verlag, https://books.google.fr/books?id=EGNQjgEACAAJ

Euclid Collaboration et al., 2019, Euclid preparation: II. The EUCLIDEMULATOR - a tool to compute the cosmology dependence of the nonlinear matter power spectrum, MNRAS, 484, 5509

Euclid Collaboration et al., 2020, Euclid preparation. VII. Forecast validation for Euclid cosmological probes, Astronomy and Astrophysics, 642, A191

Fabbian G., Calabrese M., Carbone C., 2018, CMB weak-lensing beyond the Born approximation: a numerical approach, J. Cosmology Astropart. Phys., 2018, 050

Fabbian G., Lewis A., Beck D., 2019, CMB lensing reconstruction biases in cross-correlation with large-scale structure probes, J. Cosmology Astropart. Phys., 2019, 057

Fidler C., Tram T., Rampf C., Crittenden R., Koyama K., Wands D., 2016, Relativistic interpretation of Newtonian simulations for cosmic structure formation, Journal of Cosmology and Astroparticle Physics, 2016, 031

Fields B. D., Olive K. A., Yeh T.-H., Young C., 2020, Big-Bang Nucleosynthesis after Planck, J. Cosmology Astropart. Phys., 2020, 010

Fleury P., Larena J., Uzan J.-P., 2017, Weak Gravitational Lensing of Finite Beams, Phys. Rev. Letter, 119, 191101

Fleury P., Larena J., Uzan J.-P., 2019a, Cosmic convergence and shear with extended sources, Phys. Rev. D, 99, 023525

Fleury P., Larena J., Uzan J.-P., 2019b, Weak lensing distortions beyond shear, Phys. Rev. D, 99, 023526

Fosalba P., Gaztanaga E., 1998, Cosmological Perturbation Theory and the Spherical Collapse model - I. Gaussian initial conditions, MNRAS, 301, 503

Friedrich O., Eifler T., 2018, Precision matrix expansion - efficient use of numerical simulations in estimating errors on cosmological parameters, MNRAS, 473, 4150

Friedrich O., et al., 2018, Density split statistics: Joint model of counts and lensing in cells, Phys. Rev. D, 98, 023508

Friedrich O., Uhlemann C., Villaescusa-Navarro F., Baldauf T., Manera M., Nishimichi T., 2019, Primordial non-Gaussianity without tails – how to measure fNL with the bulk of the density PDF, arXiv e-prints, p. arXiv:1912.06621

Friedrich O., et al., 2020, Dark Energy Survey Year 3 Results: Covariance Modelling and its Impact on Parameter Estimation and Quality of Fit, arXiv e-prints, p. arXiv:2012.08568

Friedrich O., Halder A., Boyle A., Uhlemann C., Britt D., Codis S., Gruen D., Hahn C., 2021, The PDF perspective on the tracer-matter connection: Lagrangian bias and non-Poissonian shot noise, arXiv e-prints, p. arXiv:2107.02300

Gavriliadis P., Athanassoulis G., 2009, Moment information for probability distributions, without solving the moment problem, II: Main-mass, tails and shape approximation, Journal of Computational and Applied Mathematics, 229, 7

Gaztañaga E., Fosalba P., Elizalde E., 2000, Gravitational Evolution of the Large-Scale Probability Density Distribution: The Edgeworth and Gamma Expansions, ApJ, 539, 522

Giocoli C., Meneghetti M., Metcalf R. B., Ettori S., Moscardini L., 2014, Mass and concentration estimates from weak and strong gravitational lensing: a systematic study, MNRAS, 440, 1899

Giocoli C., Baldi M., Moscardini L., 2018, Weak lensing light-cones in modified gravity simulations with and without massive neutrinos, MNRAS, 481, 2813

Gouin C., et al., 2019, Weak lensing in the Horizon-AGN simulation lightcone. Small-scale baryonic effects, Astronomy and Astrophysics, 626, A72

Gruen D., et al., 2018, Density split statistics: Cosmological constraints from counts and lensing in cells in DES Y1 and SDSS data, Phys. Rev. D, 98, 023507

Guth A. H., 1981, Inflationary universe: A possible solution to the horizon and flatness problems, Physical Review D, 23, 347

Halder A., Friedrich O., Seitz S., Varga T. N., 2021, The integrated three-point correlation function of cosmic shear, MNRAS, 506, 2780

Hamana T., Colombi S. T., Thion A., Devriendt J. E. G. T., Mellier Y., Bernardeau F., 2002, Source-lens clustering effects on the skewness of the lensing convergence, MNRAS, 330, 365

Harnois-Déraps J., Martinet N., Reischke R., 2021, Cosmic shear beyond 2-point statistics: Accounting for galaxy intrinsic alignment with projected tidal fields, arXiv e-prints, p. arXiv:2107.08041

Hartlap J., Simon P., Schneider P., 2006, Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix, Astronomy & Astrophysics, 464, 399–404

Hartlap J., Hilbert S., Schneider P., Hildebrandt H., 2011, A bias in cosmic shear from galaxy selection: results from ray-tracing simulations, Astronomy and Astrophysics, 528, A51

Heavens A., 2003, 3D weak lensing, MNRAS, 343, 1327

Heymans C., et al., 2021, KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints, Astronomy and Astrophysics, 646, A140

Hikage C., et al., 2019, Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data, Publications of the Astronomical Society of Japan, 71

Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, Ray-tracing through the Millennium Simulation: Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing, Astronomy and Astrophysics, 499, 31

Hilbert S., Hartlap J., Schneider P., 2011, Cosmic shear covariance: the log-normal approximation, Astronomy and Astrophysics, 536, A85

Hilbert S., et al., 2020, The accuracy of weak lensing simulations, MNRAS, 493, 305

Hu W., 1999, Power Spectrum Tomography with Weak Lensing, Astrophysical journal, 522, L21

Hui L., Gaztañaga E., Loverde M., 2007, Anisotropic magnification distortion of the 3D galaxy correlation. I. Real space, Physical Review D, 76, 103502

Hurtado-Gil L., Martínez V. J., Arnalte-Mur P., Pons-Bordería M.-J., Pareja-Flores C., Paredes S., 2017, The best fit for the observed galaxy counts-in-cell distribution function, Astronomy and Astrophysics, 601, A40

Huterer D., White M., 2005, Nulling tomography with weak gravitational lensing, Phys. Rev. D, 72, 043002

Ilbert O., et al., 2006, Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey, Astronomy and Astrophysics, 457, 841

Ivanov M. M., Kaurov A. A., Sibiryakov S., 2019, Non-perturbative probability distribution function for cosmological counts in cells, J. Cosmology Astropart. Phys., 2019, 009

Ivezić Ž., et al., 2019, LSST: From Science Drivers to Reference Design and Anticipated Data Products, ApJ, 873, 111

Jackson J. C., 1972, A critique of Rees's theory of primordial gravitational radiation, MNRAS, 156, 1P

Jalilvand M., Ghosh B., Majerotto E., Bose B., Durrer R., Kunz M., 2020, Nonlinear contributions to angular power spectra, Physical Review D, 101, 043530

Joachimi B., Schneider P., 2008, The removal of shear-ellipticity correlations from the cosmic shear signal via nulling techniques, Astronomy and Astrophysics, 488, 829

Kacprzak T., et al., 2016, Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data, MNRAS, 463, 3653

Kaiser N., 1987, Clustering in real space and in redshift space, MNRAS, 227, 1

Kaiser N., 1995, Nonlinear Cluster Lens Reconstruction, Astrophysical Journal, Letters, 439, L1

Kaiser N., Squires G., 1993, Mapping the Dark Matter with Weak Gravitational Lensing, ApJ, 404, 441

Kayo I., Taruya A., Suto Y., 2001, Probability Distribution Function of Cosmological Density Fluctuations from a Gaussian Initial Condition: Comparison of One-Point and Two-Point Lognormal Model Predictions with N-Body Simulations, ApJ, 561, 22

Kilbinger M., 2015, Cosmology with cosmic shear observations: a review, Reports on Progress in Physics, 78, 086901

Kim J., Park C., L'Huillier B., Hong S. E., 2015, Horizon Run 4 Simulation: Coupled Evolution of Galaxies and Large-Scale Structures of the Universe, Journal of Korean Astronomical Society, 48, 213

Klypin A., Prada F., Betancort-Rijo J., Albareti F. D., 2018, Density distribution of the cosmological matter field, MNRAS, 481, 4588

Krause E., Chang T.-C., Doré O., Umetsu K., 2013, The Weight of Emptiness: The Gravitational Lensing Signal of Stacked Voids, Astrophysical Journal, Letters, 762, L20

Laureijs R., et al., 2011, Euclid Definition Study Report, arXiv e-prints, p. arXiv:1110.3193

Leicht O., Uhlemann C., Villaescusa-Navarro F., Codis S., Hernquist L., Genel S., 2019, Extreme spheres: counts-in-cells for 21cm intensity mapping, MNRAS, 484, 269

Lemos P., Challinor A., Efstathiou G., 2017, The effect of Limber and flat-sky approximations on galaxy weak lensing, J. Cosmology Astropart. Phys., 2017, 014

Lewis A., Bridle S., 2002, Cosmological parameters from CMB and other data: A Monte Carlo approach, Phys. Rev. D, 66, 103511

Liu J., Madhavacheril M. S., 2019, Constraining neutrino mass with the tomographic weak lensing one-point probability distribution function and power spectrum, Phys. Rev. D, 99, 083508

Liu J., Bird S., Matilla J. M. Z., Hill J. C., Haiman Z., Madhavacheril M. S., Petri A., Spergel D. N., 2018, MassiveNuS: cosmological massive neutrino simulations, Journal of Cosmology and Astroparticle Physics, 2018, 049–049

LoVerde M., 2014, Spherical collapse in $\nu\Lambda$CDM, Physical Review D, 90, 083518

LoVerde M., Afshordi N., 2008, Extended Limber approximation, Phys. Rev. D, 78, 123506

Marozzi G., Fanizza G., Di Dio E., Durrer R., 2018, CMB lensing beyond the leading order: Temperature and polarization anisotropies, Phys. Rev. D, 98, 023535

Marques G. A., Liu J., Zorrilla Matilla J. M., Haiman Z., Bernui A., Novaes C. P., 2019, Constraining neutrino mass with weak lensing Minkowski Functionals, J. Cosmology Astropart. Phys., 2019, 019

Martinet N., Harnois-Déraps J., Jullo E., Schneider P., 2020, Probing dark energy with tomographic weak-lensing aperture mass statistics, arXiv e-prints, p. arXiv:2010.07376

Matilla J. M. Z., Waterval S., Haiman Z., 2020, Optimizing Simulation Parameters for Weak Lensing Analyses Involving Non-Gaussian Observables, The Astronomical Journal, 159, 284

Mead A., Brieden S., Tröster T., Heymans C., 2020, HMcode-2020: Improved modelling of non-linear cosmological power spectra with baryonic feedback, arXiv e-prints, p. arXiv:2009.01858

Mukhanov V., 2005, Physical Foundations of Cosmology. Cambridge Univ. Press, Cambridge, doi:10.1017/CBO9780511790553, `https://cds.cern.ch/record/991646`

Naab T., Ostriker J. P., 2017, Theoretical Challenges in Galaxy Formation, Annual Review of Astronomy and Astrophysics, 55, 59

Namikawa T., Bouchet F. R., Taruya A., 2018, CMB lensing bispectrum as a probe of modified gravity theories, Phys. Rev. D, 98, 043530

Namikawa T., Bose B., Bouchet F. R., Takahashi R., Taruya A., 2019, CMB lensing bispectrum: Assessing analytical predictions against full-sky lensing simulations, Phys. Rev. D, 99, 063511

Pajer E., van der Woude D., 2018, Divergence of perturbation theory in large scale structures, J. Cosmology Astropart. Phys., 2018, 039

Patton K., Blazek J., Honscheid K., Huff E., Melchior P., Ross A. J., Suchyta E., 2017, Cosmological constraints from the convergence 1-point probability distribution, Monthly Notices of the Royal Astronomical Society, 472, 439

Peebles P. J. E., 1980, The large-scale structure of the universe. Princeton University Press

Peel A., Pettorino V., Giocoli C., Starck J.-L., Baldi M., 2018, Breaking degeneracies in modified gravity with higher (than 2nd) order weak-lensing statistics, Astronomy and Astrophysics, 619, A38

Percival W. J., White M., 2009, Testing cosmological structure formation using redshift-space distortions, MNRAS, 393, 297

Percival W. J., et al., 2014, The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: including covariance matrix errors, MNRAS, 439, 2531

Perlmutter S., et al., 1999, Measurements of Ω and Λ from 42 High-Redshift Supernovae, The Astrophysical journal, 517, 565

Petri A., Haiman Z., Hui L., May M., Kratochvil J. M., 2013, Cosmology with Minkowski functionals and moments of the weak lensing convergence field, Phys. Rev. D, 88, 123002

Petri A., Haiman Z., May M., 2016a, Sample variance in weak lensing: How many simulations are required?, Phys. Rev. D, 93, 063524

Petri A., May M., Haiman Z., 2016b, Cosmology with photometric weak lensing surveys: Constraints with redshift tomography of convergence peaks and moments, Phys. Rev. D, 94, 063534

Petri A., Haiman Z., May M., 2017, Validity of the Born approximation for beyond Gaussian weak lensing observables, Phys. Rev. D, 95, 123503

Planck Collaboration et al., 2020a, Planck 2018 results. I. Overview and the cosmological legacy of Planck, Astronomy and Astrophysics, 641, A1

Planck Collaboration et al., 2020b, Planck 2018 results. VI. Cosmological parameters, Astronomy & Astrophysics, 641, A6

Pratten G., Lewis A., 2016, Impact of post-Born lensing on the CMB, J. Cosmology Astropart. Phys., 2016, 047

Raccanelli A., et al., 2013, Testing gravity using large-scale redshift-space distortions, MNRAS, 436, 89

Reimberg P., Bernardeau F., 2018, Large deviation principle at work: Computation of the statistical properties of the exact one-point aperture mass, Phys. Rev. D, 97, 023524

Repp A., Szapudi I., 2018, Precision prediction for the cosmological density distribution, MNRAS, 473, 3598

Riess A. G., et al., 1998, Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant, The Astronomical Journal, 116, 1009

Riess A. G., et al., 2018, Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant, The Astrophysical Journal, 861, 126

Rota G.-C., 2001, Twelve problems in probability no one likes to bring up. Springer Milan, Milano, pp 57–93, https://doi.org/10.1007/978-88-470-2107-5_5

Sachs R., 1961, Gravitational Waves in General Relativity. VI. The Outgoing Radiation Condition, Proceedings of the Royal Society of London Series A, 264, 309

Sachs R. K., Wolfe A. M., 1967, Perturbations of a Cosmological Model and Angular Variations of the Microwave Background, The Astrophysical Journal, 147, 73

Samuroff S., Mandelbaum R., Blazek J., 2021, Advances in Constraining Intrinsic Alignment Models with Hydrodynamic Simulations, MNRAS,

Schäfer B. M., Heisenberg L., Kalovidouris A. F., Bacon D. J., 2012, On the validity of the Born approximation for weak cosmic flexions, MNRAS, 420, 455

Schneider P., 1996, Detection of (dark) matter concentrations via weak gravitational lensing, MNRAS, 283, 837

Schneider P., Seitz C., 1995, Steps towards nonlinear cluster inversion through gravitational distortions. I. Basic considerations and circular clusters., Astronomy and Astrophysics, 294, 411

Schneider P., van Waerbeke L., Jain B., Kruse G., 1998, A new measure for cosmic shear, MNRAS, 296, 873

Schneider P., van Waerbeke L., Mellier Y., 2002, B-modes in cosmic shear from source redshift clustering, Astronomy and Astrophysics, 389, 729

Schneider et al., 2019a, Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation, JCAP, 2019, 020

Schneider A., Teyssier R., Stadel J., Chisari N. E., Le Brun A. M. C., Amara A., Refregier A., 2019b, Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation, J. Cosmology Astropart. Phys., 2019, 020

Scoccimarro R., Couchman H. M. P., Frieman J. A., 1999, The Bispectrum as a Signature of Gravitational Instability in Redshift Space, Astrophysical Journal, 517, 531

Seitz C., Schneider P., 1997, Steps towards nonlinear cluster inversion through gravitational distortions. III. Including a redshift distribution of the sources., Astronomy and Astrophysics, 318, 687

Sellentin E., Heavens A. F., 2016, Parameter inference with estimated covariance matrices, MNRAS, 456, L132

Shan H., et al., 2018, KiDS-450: cosmological constraints from weak lensing peak statistics - I. Inference from analytical prediction of high signal-to-noise ratio convergence peaks, MNRAS, 474, 1116

Shapiro C., Cooray A., 2006, The Born and lens lens corrections to weak gravitational lensing angular power spectra, J. Cosmology Astropart. Phys., 2006, 007

Shin J., Kim J., Pichon C., Jeong D., Park C., 2017, New Fitting Formula for Cosmic Nonlinear Density Distribution, ApJ, 843, 73

Slepian Z., Portillo S. K. N., 2018, Too hot to handle? Analytic solutions for massive neutrino or warm dark matter cosmologies, Monthly Notices of the Royal Astronomical Society, 478, 516–529

Springel V., 2010, E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh, MNRAS, 401, 791

Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, Revising the Halofit Model for the Nonlinear Matter Power Spectrum, ApJ, 761, 152

Takahashi R., Hamana T., Shirasaki M., Namikawa T., Nishimichi T., Osato K., Shiroyama K., 2017, Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments, ApJ, 850, 24

Taruya A., Takada M., Hamana T., Kayo I., Futamase T., 2002, Lognormal Property of Weak-Lensing Fields, ApJ, 571, 638

Taruya A., Bernardeau F., Nishimichi T., Codis S., 2012, Direct and fast calculation of regularized cosmological power spectrum at two-loop order, Physical Review D, 86, 103528

Taylor P. L., Bernardeau F., Kitching T. D., 2018, k -cut cosmic shear: Tunable power spectrum sensitivity to test gravity, Phys. Rev. D, 98, 083514

Taylor P. L., Bernardeau F., Huff E., 2021, x -cut Cosmic shear: Optimally removing sensitivity to baryonic and nonlinear physics with an application to the Dark Energy Survey year 1 shear data, Phys. Rev. D, 103, 043531

Teyssier R., 2002, Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES, Astronomy and Astrophysics, 385, 337

The Dark Energy Survey Collaboration 2005, The Dark Energy Survey, arXiv e-prints, pp astro–ph/0510346

Thiele L., Hill J. C., Smith K. M., 2020, Accurate analytic model for the weak lensing convergence one-point probability distribution function and its autocovariance, Phys. Rev. D, 102, 123545

Tisserand P., et al., 2007, Limits on the Macho content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds, Astronomy and Astrophysics, 469, 387

Touchette H., 2009, The large deviation approach to statistical mechanics, Physics Reports, 478, 1

Tröster T., Van Waerbeke L., 2014, Weak lensing corrections to tSZ-lensing cross correlation, J. Cosmology Astropart. Phys., 2014, 008

Troxel M. A., Ishak M., 2015, The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology, Physics Report, 558, 1

Uhlemann C., Codis S., Pichon C., Bernardeau F., Reimberg P., 2016, Back in the saddle: large-deviation statistics of the cosmic log-density field, MNRAS, 460, 1529

Uhlemann C., Codis S., Kim J., Pichon C., Bernardeau F., Pogosyan D., Park C., L'Huillier B., 2017a, Beyond Kaiser bias: mildly non-linear two-point statistics of densities in distant spheres, MNRAS, 466, 2067

Uhlemann C., Codis S., Hahn O., Pichon C., Bernardeau F., 2017b, Two is better than one: joint statistics of density and velocity in concentric spheres as a cosmological probe, MNRAS, 469, 2481

Uhlemann C., et al., 2018a, A question of separation: disentangling tracer bias and gravitational non-linearity with counts-in-cells statistics, MNRAS, 473, 5098

Uhlemann C., Pajer E., Pichon C., Nishimichi T., Codis S., Bernardeau F., 2018b, Hunting high and low: disentangling primordial and late-time non-Gaussianity with cosmic densities in spheres, MNRAS, 474, 2853

Uhlemann C., Pichon C., Codis S., L'Huillier B., Kim J., Bernardeau F., Park C., Prunet S., 2018c, Cylinders out of a top hat: counts-in-cells for projected densities, MNRAS, 477, 2772

Uhlemann C., Friedrich O., Villaescusa-Navarro F., Banerjee A., Codis S. r., 2019, Fisher for complements: Extracting cosmology and neutrino mass from the counts-in-cells PDF, arXiv e-prints, p. arXiv:1911.11158

Valageas P., 2002, Dynamics of gravitational clustering. II. Steepest-descent method for the quasi-linear regime, Astronomy and Astrophysics, 382, 412

Vallis Z. M., Wallis C. G. R., Kitching T. D., 2018, On the effect of projections on convergence peak counts and Minkowski functionals, Astronomy and Computing, 24, 84

Vicinanza M., Cardone V. F., Maoli R., Scaramella R., Er X., 2018, Increasing the lensing figure of merit through higher order convergence moments, Phys. Rev. D, 97, 023519

Villaescusa-Navarro F., et al., 2020, The Quijote Simulations, The Astrophysical Journal Supplement Series, 250, 2

Weiss A. J., Schneider A., Sgier R., Kacprzak T., Amara A., Refregier A., 2019, Effects of baryons on weak lensing peak statistics, arXiv e-prints, p. arXiv:1905.11636

Wolfram R., 2021, Mathematica, Version 12.3.1, https://www.wolfram.com/mathematica

Wong K. C., et al., 2020, H0LiCOW – XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes, MNRAS, 498, 1420

Xavier H. S., Abdalla F. B., Joachimi B., 2016, Improving lognormal models for cosmological fields, MNRAS, 459, 3693

Yang T., Birrer S., Hu B., 2020, The first simultaneous measurement of Hubble constant and post-Newtonian parameter from time-delay strong lensing, MNRAS, 497, L56

Zorrilla Matilla J. M., Sharma M., Hsu D., Haiman Z., 2020, Interpreting deep learning models for weak lensing, arXiv e-prints, p. arXiv:2007.06529

Zürcher D., Fluri J., Sgier R., Kacprzak T., Refregier A., 2021, Cosmological forecast for non-Gaussian statistics in large-scale weak lensing surveys, Journal of Cosmology and Astroparticle Physics, 2021, 028

de Lapparent V., Geller M. J., Huchra J. P., 1986, A Slice of the Universe, The Astrophysical Journal Letter, 302, L1

Planck Collaboration 2020, Planck 2018 results, Astronomy & Astrophysics, 641, A6

Alexandre BARTHELEMY **14 décembre 2021**

Sujet : "Modélisation théorique d'effets non-linéaires sur la statistique des champs faibles d'astigmatisme cosmique"

Résumé : L'astigmatisme cosmique est une conséquence directe de la relativité générale d'Einstein (ou de ses extensions) et décrit comment la présence de matière – au sens large du tenseur énergie-impulsion – déforme les trajectoires des rayons lumineux se propageant autrement en ligne droite. Cet effet, faible lorsqu'appliqué aux très grandes structures de l'Univers, se traduit par de très légères déformations des formes des galaxies d'arrière plan qui sont ainsi vues comme des variables aléatoires très corrélées puisque pour partie résultantes des corrélations existant dans les structures du champ de matière lui-même. L'étude de ces corrélations du champ des formes des galaxies est ainsi une sonde de la structuration du champ de matière totale, structuration qui contient en elle de fortes signatures, entre autres, du modèle cosmologique et de la théorie de la gravitation s'approchant au mieux des lois naturelles à l'oeuvre dans notre Univers. Malheureusement, les équations que nous utilisons pour décrire la formation des grandes structures de notre Univers n'ont pour le moment pas de solutions analytiques ce qui complique notre étude. Ainsi, en complément des outils numériques utilisés par une partie de la communauté scientifique, ce travail de thèse s'attache à proposer une modélisation théorique de la statistique à un point de certaines composantes du champs de déformation des galaxies: la convergence et la masse d'ouverture, toutes deux rattachées à la déformation isotrope des galaxies d'arrière plan. En utilisant une approche inspirée de la théorie mathématique des grandes déviations et une ré-organisation de l'information connue sous le nom de transformation de BNT, sont construites et proposées de nouvelles observables riches en information cosmologique et accessibles par une modélisation purement théorique partant de premiers-principes communément admis dans la communauté scientifique en Astrophysique et Cosmologie.

Mots clés : Cosmologie, théorie, grandes structures de l'univers, astigmatisme cosmique

Subject : "Theoretical modelling of non-linear effects on the statistics of weak-gravitational-lensing fields"

Abstract: Gravitational lensing is a direct consequence of Einstein's theory of general relativity and its extensions. It describes how the presence of matter – or more generally of a stress–energy tensor – changes how light propagates through the Universe. When applying this effect to the very large-scale structures of our Universe, we find that the images of distant galaxies that we observe are not exactly accurate but instead are slightly distorted because of this effect. This results in galaxy shapes becoming correlated since now partly resulting from the correlation of the matter field itself which is responsible for the observed deformation. The study of these correlations is thus a direct probe of the cosmological model and the theory of gravitation that mimic best the natural laws at play in our Universe. Unfortunately, the equations that we use to describe the formation of large structures in our universe do not admit any analytical solutions which lead part of the astrophysics community to rely on numerical simulations to gain information on how our models influence the formation of structures. My thesis work is complementary to these approaches as it proposes a pure theoretical modelling of certain components of the one-point statistics of the deformation field, namely the convergence and the aperture mass that are both related to the isotropic deformation of the image of background sources. Using an approach inspired by the mathematical theory of large deviations and a re-organisation of the contributing matter-structures along the lines of sight known as BNT transform, I build and propose an accurate theoretical model for cosmological-information-rich observables based on first physical principles commonly admitted in the community.

Keywords : Cosmology, theory, large-scale structure of our universe, weak gravitational lensing