Service d'Astrophysique, CEA Saclay
École doctorale Astronomie et Astrophysique d'Île-de-France
Université Paris Diderot 7 - Sorbonne Paris Cité

PhD Thesis in Astrophysics

Cosmology with weak-lensing peak counts

Chieh-An LIN

September 28th, 2016	Defense date
Advisor	Martin KILBINGER
Co-advisor	Marc SAUVAGE
Simona MEI	Jury president
David BACON	Referee
Cristiano PORCIANI	Referee
Matteo MATURI	Examinator
Sophie MAUROGORDATO	Examinator

To those who struggle to find a balance between
life, career, and curiosity

Acknowledgements

My special thanks goes to François for two main reasons. You were my officemate for two years. When you work, just like me, you are incredibly noisy. I have no idea how Ming and Paniez could ignore our crappy chats and silly ideas. I am also impressed by your sense of humor, geekness, diligence, and, most notably your perpetual willingness to help others with their problems. Whether it's about cosmology, signal processing, coding, or even LaTeX, you never say no and would always have come up with some brilliant solutions. I know if I ask you to draw me a sheep immediately without asking why (Saint-Exupéry 1943). Your helps concerning various subjects are considerable.

My deepest thanks goes to my advisor Martin. It is a truth universally acknowledged, that a naive student in possession of no ideas must be in want of a good advisor, and I am extremely fortunate to have a great one (Austen 1813). You let me interrupt you with my up-to-hour discussions on all kinds of topics. You helped me learn to bring initiatives to projects and build collaborations. You encouraged me to travel as much as I want. I will never forget the investigation of our common academic ancestors in Heidelberg and the PhD Comics you posted on my office door. You are friendly and considerate; you think with a scientific mind and a sense of humor. No one can ask for more.

Abstract

English version

Weak gravitational lensing (WL) causes distortions of galaxy images and probes massive structures on large scales, allowing us to understand the late-time evolution of the Universe. One way to extract the cosmological information from WL is to use peak statistics. Peaks are tracers of massive halos and therefore probe the mass function. They retain non-Gaussian information and have already been shown as a promising tool to constrain cosmology. In this work, we develop a new model to predict WL peak counts. The model generates fast simulations based on halo sampling and selects peaks from the derived lensing maps. This approach has three main advantages. First, the model is very fast: only several seconds are required to perform a realization. Second, including realistic conditions is straightforward. Third, the model provides the full distribution information because of its stochasticity. We show that our model agrees well with N-body simulations. Then, we study the impacts of the cosmology-dependent covariance on constraints and explore different parameter inference methods. A special focus is put on approximate Bayesian computation (ABC), an accept-reject sampler without the need to estimate the likelihood. We show that ABC is able to yield robust constraints with much reduced time costs. Several filtering techniques are studied to improve the extraction of multiscale information. Finally, the new model is applied to the CFHTLenS, KiDS DR1/2, and DES SV data sets. Our preliminary results agree with the Planck constraints assuming the Lambda-CDM model. Overall, the thesis forges an innovative tool for future WL surveys.

French version

L'effet de lentille gravitionnelle faible (WL) déforme les images des galaxies observées. Il porte des informations sur les grandes structures et nous apprend l'évolution de l'Univers. Une façon d'extraire des informations de WL est d'utiliser les statistiques d'ordres supérieurs à deux, en particulier le comptage de pics. Les pics indiquent la présence des halos et mesurent la fonction de masse. Ils sont considérés comme un bon outil pour contraindre la cosmologie. Dans cette thèse, on développe un nouveau modèle de prédiction sur le nombre des pics. Celui-ci génère des simulations des halos et sélectionne les pics à partir de la carte de WL résultante. Cette approche jouit de trois avantages. D'abord, le modèle est rapide: une réalisation s'obtient en quelques secondes. Ensuite, il est aisé d'y inclure les effets observationnels. Enfin, on a accès à la distribution complète des observables grâce à son caractère stochastique. On montre que notre modèle est en accord avec les simulations à N-corps. Puis, on examine des méthodes de contraintes variées. Un accent est mis sur la computation bayesienne approximative (ABC) qui est un processus d'acceptation/rejet ne nécessitant pas d'évaluer la fonction de vraisemblance. On montre que ABC fournit des contraintes cohérentes en un temps plus

faible que la méthode classique. Des méthodes de filtrage, pour améliorer l'extraction des informations multi-échelles, sont étudiées. Enfin, le nouveau modèle est appliqué sur les données des relevés. Nos résultats préliminaires sont en accord avec ceux de Planck, en admettant le modèle Lambda-CDM. Dans l'ensemble, cette thèse bâtit un chemin pionier pour les futurs relevés de WL.

Contents

Chapter I

Introduction

Measure what is measurable, and make measurable what is not so.

— Galileo Galilei

Since existence, human beings have never stopped their curiosity about the most outer frontier of the world. From bare-eye observations to telescopes, cosmology — the science of the Universe — has been developed little by little. What makes up the Universe? What is the history of it? These might be the two ultimate questions that all cosmologists try to answer. So far, we estimate that the age of the Universe is roughly 13.8 billion years, and that up to 95% is actually composed of "dark components" (Planck Collaboration et al. 2015a): with *dark matter*, we manage to explain the luminous mass deficit from observations; and *dark energy* is responsible for the accelerating cosmic expansion. Only about 5% of the Universe is ordinary matter: baryons in the form of ions, atoms, and molecules.

During the last twenty years, the probes of the cosmic microwave background confirm that our Universe can actually be described by a simple model with only six parameters (Spergel et al. 2003), namely "ΛCDM". This success, both for the probes and for the model, defines an important benchmark in cosmology. Until now, almost all observations are consistent with the ΛCDM model. For this reason, cosmologists aim more and more for improving the uncertainty of the measurements. We have been passing through the era of "precision cosmology" for a while. However, several recent studies (MacCrann et al. 2015; Planck Collaboration et al. 2015b) reveal the existence of tensions in the ΛCDM model, which opens a door for possible new physics in the future.

To verify if new physics exists, one of the viable methods is gravitational lensing, or lensing in short. The origin of lensing is the light deflection by gravity: when light travels in the Universe, the presence of massive structures perturb the local space-time curvature, so that the straight light path is deflected. This phenomenon has been well predicted and modelled by the famous theory of general relativity (Einstein 1915). Due to this deflection, distant sources yield distorted images when the emitted light reaches observers. This distortion encodes cumulatively the gravitational information along the line of sight. Therefore, by measuring properly the image distortion, cosmologists can measure the distribution of mass in the Universe.

When cosmologists observe multiple images of the same source, or ring- or arc-like patterns, since such objects could not exist in the Universe, they know that images have been

15

strongly lensed. In this case, we call it "strong lensing". On the contrary, the lensing effect is so weak for most parts of the sky such that distinguishing lensed and unlensed images by eye is almost impossible. In this case, we call it "weak lensing". Despite its low signal level, weak lensing still contains very rich cosmological information.

To extract cosmological information from weak lensing, a very common way is to use second-order statistics. However, this only retains the Gaussian information (Gauss 1809), and it misses the rich nonlinear part of structure evolution encoded on small scales. To compensate for this drawback, several non-Gaussian statistics have been proposed, for example higher order moments, the three-point correlation function, Minkowski functionals, or peak statistics.

The non-Gaussian observable that is studied in this thesis is peak counts. Weak-lensing peaks are defined as the local maxima on a convergence map. Since the weak-lensing convergence represents the projected mass, peaks are tracers of massive regions in the Universe, which are characterized by the mass function. For this reason, we probe the mass function and constrain the cosmology simply by counting peaks.

However, not all the peaks are true mass clusters. The galaxy shape noise as well as real-world observing conditions create selection effects, such as false positives, false negatives, or changes of peak height. In the literature, we see that early studies tended to use only very high signal-to-noise ratio ($\gtrsim 5$) peaks. In this way, cosmologists can guarantee that all peaks are true clusters and proceed with the physical analysis with identifications. However, recent studies show that including selection effects directly into the model allows us to explore medium and low peaks, which increases cosmological information extraction.

So far, there exist three approaches which predict peak counts by taking selection effects into account: analytical modelling, N-body simulations, and a fast stochastic model (this work). Analytical models are based on Gaussian random field theory and peak theory. This approach is not very flexible. It suffers from real-world effects and can be biased. On the other hand, N-body simulations can include observational conditions into forward computations. However, they are very expensive to run and evaluate. That is how the motivation for a new model emerges: we want to benefit from the advantages of N-body modelling, find a shortcut to avoid complex N-body processes, and yield a correct peak-count prediction.

Nowadays, statistical advances are changing the perspectives of cosmology. The applications of sparse methods, Bayesian analysis, machine learning, etc. to cosmology has been prosper. As the volume of data and the requirement of the precision on results grow, using sophisticated statistical tools becomes indispensable for cosmological studies.

Recently, approximate Bayesian computation (ABC) has been gathering a strong growth of attention from the astrophysical community. ABC is a likelihood-free parameter inference method. Usually, it is used when the likelihood is expensive to evaluate or impossible to define. However, even if the likelihood can be computed in a straightforward way, ABC is still powerful and it has shown a great potential to provide cosmological constraints.

The objective of this thesis is to address the following questions:

- How to model weak-lensing peak counts properly in realistic conditions?

- Which parameter constraint method is preferred?

- How to compare different filtering techniques?

- What cosmological information can we extract from peaks?

The content of the thesis is structured as follows. In Chaps. 2, 3, and 4, some concepts of cosmology, structure formation, and weak gravitational lensing that are used in this work will be successively presented. A review of the weak-lensing-peak studies in the literature will be given at the end of Chap. 4. In Chap. 5, I will introduce the problem of peak-count modelling and propose a solution to overcome it. A comparison to N-body simulation and to the model from Fan et al. (2010) will be provided. In Chap. 6, I will discuss the results from different constraint strategies. A particular attention will be put on the cosmology-dependent-covariance effect. In Chap. 7, approximate Bayesian computation will be introduced and the resulting cosmological constraints will be shown. In Chap. 8, I will compare different filtering techniques, including linear and nonlinear ones. In Chap. 9, the new peak-count model and ABC will be applied to data retrieved from three different lensing surveys. Finally, I will summarize the discoveries and the results of this thesis in Chap. 10.

Chapter II

Modern cosmology

Overview

This chapter will introduce the cosmological basis for the thesis, particularly Friedmann's equations, cosmological distances, and a brief description of the history of the Universe.

2.1 Cosmic expansion history

2.1.1 ΛCDM model

From the current observations, the Universe is very well described by the ΛCDM model, where Λ stands for the cosmological constant, the simplest parametrization of "dark energy" and CDM stands for cold "dark matter".

The concept of *dark matter* was first proposed by Zwicky (1933) to describe the mass deficit from luminosity compared to the gravitational mass necessary for galaxy random movement in cluster virial theorem. Even by considering non-luminous baryons, the deficit can still not be fully explained because the baryonic fraction is well constrained by Big-Bang nucleosynthesis and baryonic acoustic oscillations. Some possible candidates of dark matter are elementary particles, for example those which are motivated by supersymmetry theory. Alternatively, this mass deficit problem can also be solved by modifying the gravity laws. However so far, no direct detection of dark matter particles has been claimed, and no modified gravity theory can match current observation results in a satisfactory way.

On the other hand, *dark energy* is a key to explain the acceleration of the expansion of the Universe, which was confirmed by Riess et al. (1998). The simplest dark energy model is the cosmological constant, which conducts a constant pressure per volume which pushes the Universe out.

The ΛCDM model contains in total six parameters. Six is the minimal number of parameters to fit the observations, so the ΛCDM model can be considered as the simplest model. Apart from the physical fractions of baryon, cold dark matter, and dark energy ($\Omega_{\mathrm{b}}h^2$, $\Omega_c h^2$, Ω_Λ), three other parameters are the amplitude and spectral index of the scalar field power spectrum ($\Delta_{\mathcal{R}}^2$, n_{s}) and the reionization optical depth τ. According to recent studies (Planck Collaboration et al. 2015a), ordinary matter only accounts for 5% of the total energy of the Universe; 26% are dark matter; and 69%, the majority, are dark energy, the nature of which

is totally unknown.

2.1.2 Einstein field equation

The current understanding of the Universe is founded on the theory of General Relativity (Einstein 1915). In the theory, time and space are related together by the *metric* $\mathrm{d}s^2 \equiv g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$ and form one single space called space-time. Then, gravity, the dominant force at cosmic scales, is described as the distortion of the space-time curvature, which explains the deviation from straight lines of the trajectories of travelling particles. For example, light (or photons) will always travel on the zero-geodesics.

Let us consider an homogeneous and isotropic universe. This ideal scenario is described by the *Friedmann-Lemaître-Robertson-Walker* (FLRW) *metric* as

$$\begin{aligned}
\mathrm{d}s^2 &= -\mathrm{c}^2\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}R^2}{1-KR^2} + R^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right] \\
&= -\mathrm{c}^2\mathrm{d}t^2 + a^2(t)\left[\mathrm{d}w^2 + f_K^2(w)\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right]
\end{aligned} \tag{2.1}$$

where c is the light speed, t is the cosmic time, $a(t)$ is the scale factor, R, θ, and ϕ are spherical coordinates in comoving space, K is the flatness of the universe which characterizes the relation between space and time, and w and $f_K(w)$ are respectively comoving radial and transverse distances (see Sect. 2.2.2). A flat universe corresponds to $K = 0$, whereas a closed (or elliptic) universe has $K > 0$ and an open (or hyperbolic) universe has $K < 0$.

The relation between the curvature distortion and cosmic components (matter, radiation, etc.) is described by the Einstein field equation. It links the space-time curvature to the energy-momentum. The version with the cosmological constant Λ is

$$G_{\mu\nu} = \frac{8\pi\mathrm{G}}{\mathrm{c}^4}T_{\mu\nu} - \Lambda g_{\mu\nu}, \tag{2.2}$$

where G is the gravitational constant, $g_{\mu\nu}$ is the metric, and $G_{\mu\nu}$ is the Einstein tensor which can be written with the Ricci tensor $R_{\mu\nu}$ and the Ricci scalar R, as

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}, \tag{2.3}$$

and $T_{\mu\nu}$ is the *stress-energy tensor*.

2.1.3 Friedmann's equations

For an isotropic homogeneous universe, one can set the $(1,1)-$order tensor $T^\mu{}_\nu$ to

$$T^\mu{}_\nu = \begin{pmatrix} -\rho(t) & 0 & 0 & 0 \\ 0 & p(t) & 0 & 0 \\ 0 & 0 & p(t) & 0 \\ 0 & 0 & 0 & p(t) \end{pmatrix}, \tag{2.4}$$

where $\rho(t)$ denotes the mass density and $p(t)$ denotes the pressure. This allows one to simplify the Einstein equation into a simpler form. The 00-component and the trace respectively lead

to

$$\frac{\dot{a}^2}{a^2} + \frac{c^2 K}{a^2} = \frac{8\pi G}{3}\rho + \frac{c^2 \Lambda}{3}, \tag{2.5}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{c^2 \Lambda}{3}, \tag{2.6}$$

where the notation $\dot{}$ is the derivative with regard to the cosmic time t. Equations (2.5) and (2.6), called *Friedmann's equations*, allow one to derive the evolutions of each cosmic component and of the whole Universe. Taking the derivative of Eq. (2.5) and using Eq. (2.6), one has

$$\begin{aligned}
\frac{8\pi G}{3}\dot{\rho} &= 2\left(\frac{\dot{a}}{a}\right)\left(\frac{a\ddot{a} - \dot{a}^2}{a^2}\right) - \frac{2c^2 K \dot{a}}{a^3} \\
&= 2\left(\frac{\dot{a}}{a}\right)\left(\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} - \frac{c^2 K}{a^2}\right) \\
&= 2\left(\frac{\dot{a}}{a}\right)\left[\left(-\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{c^2 \Lambda}{3}\right) - \left(\frac{8\pi G}{3}\rho + \frac{c^2 \Lambda}{3}\right)\right] \\
&= \frac{8\pi G}{3}\left(\frac{\dot{a}}{a}\right)\left(-\rho - \frac{3p}{c^2} - 2\rho\right),
\end{aligned} \tag{2.7}$$

so

$$\dot{\rho} = -3\left(\rho + \frac{p}{c^2}\right)\frac{\dot{a}}{a}. \tag{2.8}$$

It is then useful to introduce the notion of the *equation of state*, which is the link between the density and the pressure. For each component α, we may write

$$\frac{p_\alpha}{c^2} = w_\alpha \rho_\alpha, \tag{2.9}$$

so that the general form of the component evolution is

$$\frac{\rho_\alpha}{\rho_{\alpha,0}} = \left(\frac{a}{a_0}\right)^{-3(1+w_\alpha)}, \tag{2.10}$$

where the subscript $_0$ means that the quantity is evaluated at the present time $t = t_0$. We set naturally $a_0 = 1$.

For non-relativistic components, such as baryon and cold dark matter, $w = 0$; for ultra-relativistic components such as photon and neutrinos, $w = 1/3$. We usually simplify these as two families: matter and radiation. For dark energy, one can imagine a scenario where $w = -1$. This actually corresponds to the case of the cosmological constant Λ. To see this, one only needs to consider a "density of cosmological constant" ρ_Λ such that $\Lambda = 8\pi G \rho_\Lambda / c^2$. By doing this, the contribution from dark energy is included in the ρ and p/c^2 prescriptions and Λ can be dropped. From Eq. (2.10), one can see that ρ_Λ is independent from the evolution of the Universe since $w = -1$, which justifies the name of "cosmological constant". However, a more complex model could be possible. For example,

$$w_{\rm de}(a) = w_0^{\rm de} + (1-a)w_a^{\rm de} \tag{2.11}$$

is a commonly considered equation of state for dark energy. The case of $w_{\rm de} < -1/3$ will be compatible with the expansion of the Universe.

Consider now an universe composed with the matter (labelled $\mathfrak{m}$), the radiation (labelled $\mathfrak{r}$), and the cosmological constant. We introduce the *Hubble parameter* to simplify the notation:

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)}, \tag{2.12}$$

which is also the relative expansion rate of the Universe. The first Friedmann's equation (Eq. 2.5) becomes

$$H^2 + \frac{c^2 K}{a^2} = \frac{8\pi G}{3}(\rho_m + \rho_r) + \frac{c^2 \Lambda}{3}. \tag{2.13}$$

Defining now the dimensionless fraction of different components **at the present time** as

$$\Omega_K \equiv \frac{-c^2 K}{a_0^2 H_0^2}, \quad \Omega_\Lambda \equiv \frac{c^2 \Lambda}{3 H_0^2}, \quad \Omega_m \equiv \frac{8\pi G}{3 H_0^2}\rho_m, \quad \Omega_r \equiv \frac{8\pi G}{3 H_0^2}\rho_r, \tag{2.14}$$

we obtain, by evaluating Eq. (2.13) at $t = t_0$, an energy balance:

$$\Omega_K + \Omega_\Lambda + \Omega_m + \Omega_r = 1. \tag{2.15}$$

Here, both cosmological constant and flatness have been treated as a sort of energy. Following this reasoning, the total density of the Universe at the present time, called the *critical density*, is

$$\rho_{\mathrm{crit}} \equiv \frac{3 H_0^2}{8\pi G} = 2.775 \cdot 10^{11} \; M_\odot h^2 / \mathrm{Mpc}^3. \tag{2.16}$$

This quantity has an unity of mass volume density. In this thesis, the mass unity will always be $M_\odot/h$, the solar mass over h; and the distance will always be Mpc/h, where h is the dimensionless Hubble parameter, defined as

$$H_0 \equiv 100 \cdot h \; \mathrm{km/s/Mpc} = 10^5 \cdot h \; \mathrm{m/s/Mpc}. \tag{2.17}$$

Given the current decomposition of the Universe, what is its evolution history? This can be derived from Eq. (2.13). Using Eqs. (2.10) and (2.14), We actually have

$$\frac{H^2}{H_0^2} = \Omega_K \left(\frac{a}{a_0}\right)^{-2} + \Omega_\Lambda + \Omega_m \left(\frac{a}{a_0}\right)^{-3} + \Omega_r \left(\frac{a}{a_0}\right)^{-4}. \tag{2.18}$$

Since $a_0 = 1$, we usually express Eq. (2.18) in redshift $z \equiv a^{-1} - 1$ (see Sect. 2.2.1 for the definition of the redshift), so that

$$\frac{H^2(z)}{H_0^2} = E^2(z), \quad \text{with} \quad E(z) = \sqrt{\Omega_\Lambda + \Omega_K(1+z)^2 + \Omega_m(1+z)^3 + \Omega_r(1+z)^4}. \tag{2.19}$$

This means that, to reconstruct the evolution, we need to measure the fraction of each component and the Hubble rate, all at the present time. Taking these quantities as external parameters, Eq. (2.19) allows one to integrate over the lookback time and trace the evolution of a (Fig. 2.1). For example, we have $a \propto t^{1/2}$ in the radiation-dominated era and $a \propto t^{2/3}$ in the matter dominated era.

In the ΛCDM model, Ω_m is composed of cold dark matter and baryons, while $\Omega_K, \Omega_r \approx 0$.
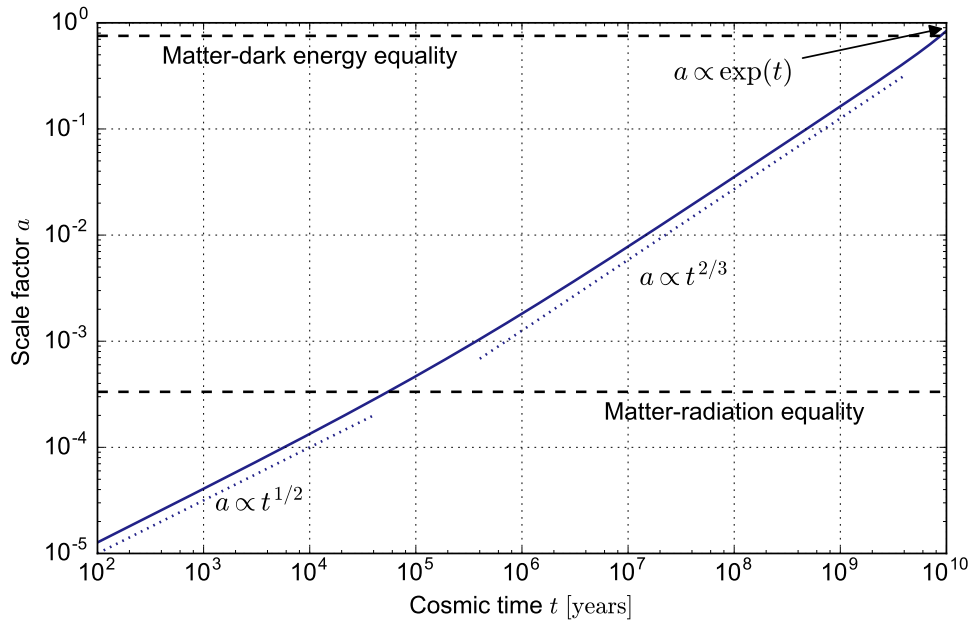
FIGURE 2.1: Evolution of the scale factor $a(t)$ with regard to the cosmic time under a ΛCDM cosmology.

2.2 Distances in the Universe

The notion of "distance" is very subtle in the cosmological context. The distances and scales that cosmologists refer to are often so large that even light takes millions of years to travel. This evokes the concept of horizon and causality. Implicitly for an observer, a distanced object is also an "old" object, since information that the observer receives was emitted long time ago. As a consequence, the information from different distances can be considered as different stages on a time sequence. Although these emissions come from different physical positions, cosmologist can still take advantage to study space-invariant properties, such as correlation functions.

2.2.1 Redshift

The time-distance duality above is formally characterized by *redshift*. The redshift is an observational phenomenon indicating that the spectrum of sources is shifted toward red sequences. This can be identified by atomic emission and absorption wavelengths. The cause of the redshift can be the Doppler effect, but also the expansion of the Universe. Actually, from the metric (Eq. 2.1), we have

$$\frac{\mathrm{c}\mathrm{d}t}{a(t)} = \mathrm{d}w \tag{2.20}$$

since $\mathrm{d}^2 s = 0$ for light. If we focus on a light ray emitted at t_e (with wavelength λ_e) and received at t_r (with λ_r), then

$$\int_{t_\mathrm{e}}^{t_\mathrm{r}} \frac{\mathrm{c}\mathrm{d}t}{a(t)} = \int_0^w \mathrm{d}w = \int_{t_\mathrm{e}+\lambda_\mathrm{e}/\mathrm{c}}^{t_\mathrm{r}+\lambda_\mathrm{r}/\mathrm{c}} \frac{\mathrm{c}\mathrm{d}t}{a(t)}, \tag{2.21}$$

which leads to

$$\int_{t_e}^{t_e + \lambda_e/c} \frac{c dt}{a(t)} = \int_{t_r}^{t_r + \lambda_r/c} \frac{c dt}{a(t)}, \tag{2.22}$$

Therefore, we can relate the wavelengths at emission and reception to their respective scale factors $a_e = a(t_e)$ and $a_r = a(t_r)$ by

$$\frac{\lambda_r}{\lambda_e} = \frac{a_r}{a_e}, \tag{2.23}$$

and the shift is defined as

$$z \equiv \frac{\lambda_r - \lambda_e}{\lambda_e}. \tag{2.24}$$

In this case, we call it cosmological redshift. This term is dominant in most cosmological contexts. In the absence of expansion, $a_r = a_e$ and the cosmological redshift disappears. Therefore, if we take $a_r = a_0 = 1$ and $a_e = a$, the redshift becomes simply

$$a = \frac{1}{1+z}. \tag{2.25}$$

As we can see from Eq. (2.18), a is a monotonic function of t if all $\Omega_i > 0$. This is the case of the current Universe. Thus, z is implicitly a one-to-one function of t. We can then parametrize the time by the redshift. At the present time, we have $z(t_0) = 0$.

2.2.2   Cosmological distances

To convert redshift to distance, we should be aware of the fact that the definition of cosmological distance is not unique. Here, we will always consider a light ray emitted at redshift z_e and received by an observer at redshift $z_r < z_e$.

Proper distance   The proper distance D_p is the distance that a photon "sees". It refers to the collection of infinitesimal light paths despite the expansion of the Universe. Mathematically, it is defined as $dD_p = c dt$. Since the light speed is a constant, the proper distance is nothing but

$$D_p(z_r, z_e) \equiv c\big(t(z_r) - t(z_e)\big). \tag{2.26}$$

Comoving radial distance   The comoving radial distance (or simply comoving distance) w is defined as $dw = ca^{-1}dt$. From the point of view of the metric, it takes the infinitesimal proper distance and rescales it by a^{-1} due to the cosmic flow. From a more global aspect, this is just the physical distance between the emission point and the reception point measured at $z = 0$. The word "comoving" always refer to a scaling to $z = 0$. Mathematically, one has $dw = ca^{-1}dt = c(a\dot{a})^{-1}da = c(a^2H)^{-1}da$, hence

$$w(z_r, z_e) \equiv D_H \int_{z_r}^{z_e} \frac{dz}{E(z)}, \tag{2.27}$$

where

$$D_H \equiv \frac{c}{H_0} = 2997.92458 \text{ Mpc}/h \tag{2.28}$$

is the Hubble distance, and $E(z)$ is given by Eq. (2.19).
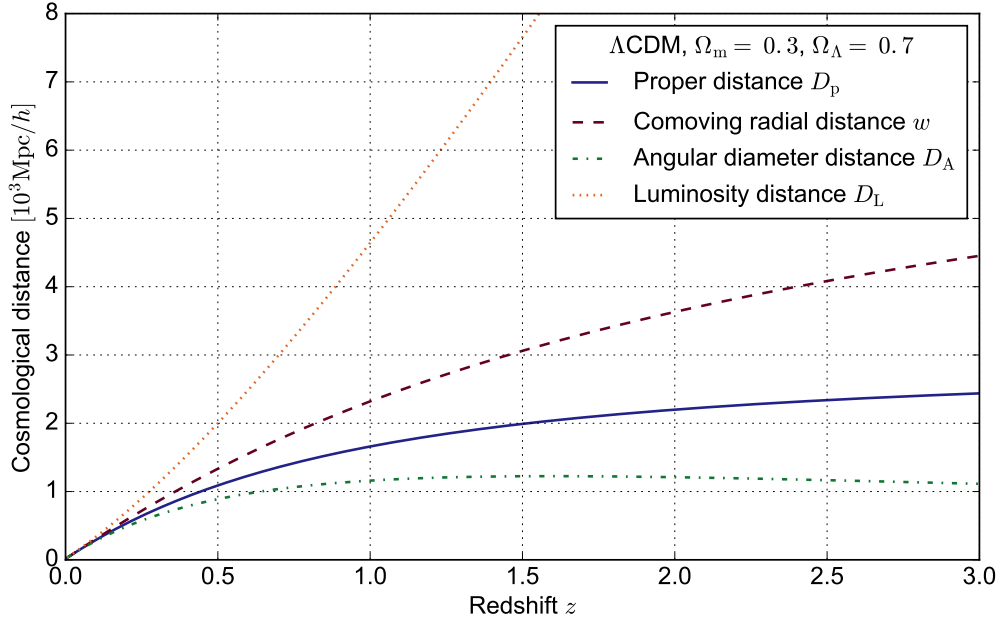
FIGURE 2.2: Different distances under a ΛCDM cosmology.

Comoving transverse distance   The comoving transverse distance f_K is the ratio of the comoving separation between two points at w to their separation angle. Once again, we scale the distance from the emission epoch z_e to $z = 0$. So, a better interpretation is to stop the cosmic flow at $z = 0$, and to measure the (physical) distance between these two distanced points. This is purely geometric, so f_K only depends on w and is related to the curvature K defined in the metric (Eq. 2.1). We have:

$$
f_K(w) \equiv \begin{cases} \dfrac{1}{\sqrt{-K}} \sinh\left(\sqrt{-K}w\right) & \text{if } K < 0, \\ w & \text{if } K = 0, \\ \dfrac{1}{\sqrt{K}} \sin\left(\sqrt{K}w\right) & \text{if } K > 0. \end{cases} \tag{2.29}
$$

Angular diameter distance   The angular diameter distance D_A is the ratio of a physical separation at the emission epoch (with comoving radial distance w) to its angular separation from an observer. This is used for calculating the size of bound objects, such as galaxies and halos, because we usually require their original size, not rescaled to $z = 0$. To recover this, we only need to scale the comoving quantity back to the emission epoch z_e. Hence, the angular diameter distance is simply

$$
D_A(z_r, z_e) \equiv \frac{f_K(w(z_r, z_e))}{1 + z_e}. \tag{2.30}
$$

Particularly, if the Universe is flat and the observer is at $z_r = 0$, the angular diameter distance becomes

$$
D_A(0, z_e) = \frac{D_H}{1 + z_e} \int_0^{z_e} \frac{\mathrm{d}z'}{E(z')}. \tag{2.31}
$$

Luminosity distance The luminosity distance D_L is defined to relate the luminosity L and the brightness B, such that

$$D_\mathrm{L} \equiv \sqrt{\frac{L}{4\pi B}}.$$ (2.32)

The dependency on redshifts is omitted. It acts only on B as the flux emitted at z_e and received at z_r. From Etherington's reciprocity theorem (Etherington 1933; Ellis 2007), we obtain a further relation:

$$D_\mathrm{A} = \frac{f_K}{1+z} = \frac{D_\mathrm{L}}{(1+z)^2}$$ (2.33)

for all observers at redshift 0 and sources at redshift z.

Comoving volume The comoving volume is simply the volume defined in comoving space. On small scales, where one can locally neglect the variation of the cosmic flow, it is the physical volume at the present time which corresponds to the physical volume of the same Lagrangian points at the considered epoch. However, if we consider a long time span, e.g. a lightcone constructed from $z = 1$ to the present time, this interpretation is not valid anymore. This difficulty reveals the fact that the comoving volume is an ill-defined concept. The problem can only be resolved if we cut the total volume into small pieces and rescale them to $z = 0$ with their respective scale factors. For these reason, in this work, except for specific precision, all volumes are physical volumes measured at the considered epoch.

2.3 Modern cosmological paradigm

Fine-tuning problems

The standard model of cosmology provided by Friedmann's equations is subject to problems evoked by two observational facts. The first is the *flatness problem*. The flatness of the Universe that we observe today is very close to zero. If we look back to the birth time, it means that the initial density of different cosmic components should be set to a specific value at extremely high precision. The second one is the *horizon problem*. It turns out that our Universe is very homogeneous on large scales. However, these scales are so large that the interactions, at most propagated at the speed of light, have not yet occurred (scales have not "entered the horizon"). One of the solutions to these two problems is the *cosmological inflation* (Guth 1981). During the inflation, the flatness would be forced very close to zero and the anisotropies would be suppressed, so that the Universe becomes what we observe now, flat and homogeneous.

A brief history of the Universe

Today, the most commonly accepted theory for the beginning of the Universe is the *Big Bang* theory, which admits a singularity in the space-time. Very soon after, cosmological inflation occurs and the primordial (quantum) perturbation signatures get dragged and remain in the Universe. The temperature drops considerably and symmetries between fundamental forces break. Successively, antibaryons annihilate, neutrinos decouple, and the *Big Bang*
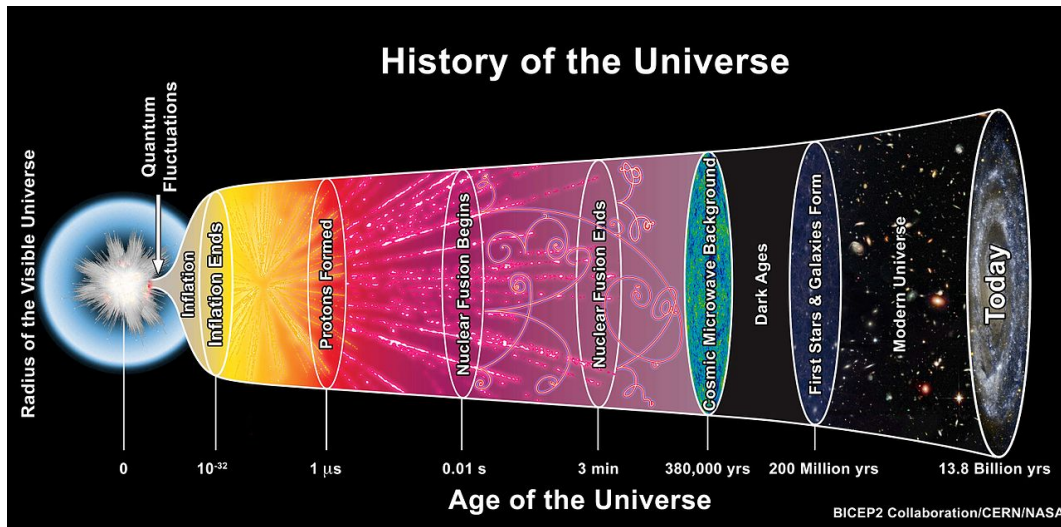
FIGURE 2.3: Illustration of the history of the Universe. (Source: BICEP2/CERN/NASA)

nucleosynthesis occurs, which forms atomic nuclei. All these processes are estimated to happen during the first few minutes of the history of the Universe.

As time goes on, the Universe continues to expand and the temperature continues to decrease. The matter-radiation equality is achieved roughly at $z \approx 3600$. Then, at $z \approx 1100$, which is about 380,000 years after the Big Bang, the temperature of the Universe falls below 4000 K. The plasma of photons, electrons, and protons is not at thermal equilibrium anymore. Electrons and protons are bound together forming hydrogen atoms, which is called *recombination*. Meanwhile, photons are decoupled from the matter and start to travel freely. This turns the Universe transparent and these photons become the cosmic microwave background (CMB) that we observe today.

After the recombination, the Universe enters into the dark age where no radiation other than CMB is emitted. During this epoch, matter collapses and structures form due to gravity. At $z \approx 20$, highly dense regions start to form stars and galaxies. These astrophysical objects emit new light, which is energetic enough to rip electrons from electroneutral atoms. This is called *reionization*. While the most distanced galaxy that humans observe so far is recorded at $z = 11.1$ (Oesch et al. 2016), the primary form of galaxy clusters only appear after $z \approx 6$. At $z \approx 0.4$, the Universe passes through the matter-dark energy equality era. Today, it is believed that the dominant component in the Universe is dark energy.

Summary

In this small chapter, we have reviewed briefly Friedmann's equations which describe the evolution of the scale factor. We have also seen that the word "distance" could have various definitions depending on the context.

The ΛCDM model is the current standard model for cosmology. With the ΛCDM model, the flatness and the horizon problems are explained with the cosmological inflation.

In next chapter, we will focus on the formation and the evolution of cosmological structures.

Chapter III

Structure formation

Overview

The goal of this chapter is to introduce two important aspects of structures in the Universe for weak-lensing peak counts: the mass function and the universal halo density profile. To have a better understanding, I will first review some key concepts such as Fourier analysis, cosmological perturbation theory, and the spherical collapse model to define matter fluctuations, the growth factor, and the linear contrast threshold. After that, different models of the mass function and the density profile will be presented.

3.1   Fourier analysis

For numerous reasons, analysis in Fourier space is often preferred than in direct space for cosmological studies. For example, the idea of horizon crossing — distant perturbations start to "communicate" with each other due to the expanding Universe — can be straightforwardly summarized as wave modes entering into the horizon. Also, randomness has been included in cosmological models to describe the observed universe (e.g. CMB temperature and large-scale structures), and an efficient way to extract cosmological information from it is to study the statistical quantities measured in the Fourier space, e.g. the power spectrum. In this section, the power spectrum, the quantity σ_8, and the formalism of the Fourier analysis will be introduced.

3.1.1   Two-point-correlation function and power spectrum

Let $f(\boldsymbol{r})$ be a 3D field defined at the real space position $\boldsymbol{r}$. The Fourier transform of $f(\boldsymbol{r})$ is defined as

$$\tilde{f}(\boldsymbol{k}) \equiv \int_{\mathbb{R}^3} \mathrm{d}^3\boldsymbol{r} \ \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\boldsymbol{r}} f(\boldsymbol{r}), \tag{3.1}$$

where $\boldsymbol{k}$ is a wave vector in the Fourier space, and the inverse transform as

$$f(\boldsymbol{r}) \equiv \int_{\mathbb{R}^3} \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \ \mathrm{e}^{\mathrm{i}\boldsymbol{k}\boldsymbol{r}} \tilde{f}(\boldsymbol{k}). \tag{3.2}$$

Consider now two fields f_1, f_2 and their convolution product:

$$(f_1 * f_2)(\boldsymbol{r}) \equiv \int \mathrm{d}^3\boldsymbol{r}' \ f_1(\boldsymbol{r} - \boldsymbol{r}')f_2(\boldsymbol{r}'). \tag{3.3}$$

This is identical to the product of f_1 and f_2 in the Fourier space since

$$\begin{aligned}
(f_1 * f_2)(\boldsymbol{r}) &= \int \mathrm{d}^3\boldsymbol{r}' \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \frac{\mathrm{d}^3\boldsymbol{k}'}{(2\pi)^3} \ \mathrm{e}^{\mathrm{i}k(r-r')}\mathrm{e}^{\mathrm{i}k'r'}\tilde{f}_1(\boldsymbol{k})\tilde{f}_2(\boldsymbol{k}') \\
&= \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \frac{\mathrm{d}^3\boldsymbol{k}'}{(2\pi)^3} \left(\int \mathrm{d}^3\boldsymbol{r}' \ \mathrm{e}^{-\mathrm{i}(k-k')r'}\right) \mathrm{e}^{\mathrm{i}kr}\tilde{f}_1(\boldsymbol{k})\tilde{f}_2(\boldsymbol{k}') \\
&= \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \frac{\mathrm{d}^3\boldsymbol{k}'}{(2\pi)^3} \ (2\pi)^3\delta^{(3)}(\boldsymbol{k} - \boldsymbol{k}')\mathrm{e}^{\mathrm{i}kr}\tilde{f}_1(\boldsymbol{k})\tilde{f}_2(\boldsymbol{k}') \\
&= \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \ \mathrm{e}^{\mathrm{i}kr}\left(\tilde{f}_1 \cdot \tilde{f}_2\right)(\boldsymbol{k}),
\end{aligned} \tag{3.4}$$

knowing that the Fourier transform of 1 is a Dirac function. In other words,

$$\widetilde{f_1 * f_2} = \tilde{f}_1 \cdot \tilde{f}_2 \tag{3.5}$$

For any 3D field f, the *two-point-correlation function* (2PCF) $C(\boldsymbol{r})$ of f is defined as

$$C(\boldsymbol{r}) = \langle f(\boldsymbol{r}')f(\boldsymbol{r}' + \boldsymbol{r})\rangle_{\boldsymbol{r}'} \equiv \int_{\mathbb{R}^3} \mathrm{d}^3\boldsymbol{r}' \ f(\boldsymbol{r}')f(\boldsymbol{r}' + \boldsymbol{r}), \tag{3.6}$$

where the subscript $_{\boldsymbol{r}'}$ stands for averaging over $\boldsymbol{r}'$. This is interpreted as the expectation value of the product of two field points which are separated by the vector $\boldsymbol{r}$. Note that this is not exactly a convolution product for the reason of sign. Actually, we have

$$C(\boldsymbol{r}) = \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \ \mathrm{e}^{\mathrm{i}kr}\tilde{f}(\boldsymbol{k})\tilde{f}(-\boldsymbol{k}) = \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \ \mathrm{e}^{\mathrm{i}kr}\left|\tilde{f}(\boldsymbol{k})\right|^2, \tag{3.7}$$

since $\tilde{f}(-\boldsymbol{k}) = \tilde{f}^*(\boldsymbol{k})$.

Let us now focus on the the contrast field and its variance. The contrast δ of a field f is given by

$$\delta(\boldsymbol{r}) \equiv \frac{f(\boldsymbol{r}) - \bar{f}}{\bar{f}} \quad \text{with} \quad \bar{f} \equiv \lim_{V \to +\infty} \frac{1}{V} \int_V \mathrm{d}^3\boldsymbol{r} \ f(\boldsymbol{r}). \tag{3.8}$$

Since the average of the contrast is zero by construction, the variance σ^2 is just the integral of δ^2 over the whole space, which leads to

$$\sigma^2 \equiv \left\langle \delta^2(\boldsymbol{r})\right\rangle_{\boldsymbol{r}} = \int_{\mathbb{R}^3} \mathrm{d}^3\boldsymbol{r} \ \delta^2(\boldsymbol{r}). \tag{3.9}$$

Note that, depending on the field, this quantity can diverge. We can see that Eq. (3.9) is equivalent to Eq. (3.7) by substituting f with δ. Keeping the same symbol for the 2PCF, σ^2 now becomes

$$\sigma^2 = C(0) = \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \ \left|\tilde{\delta}(\boldsymbol{k})\right|^2 = \int \frac{k^2\mathrm{d}k \ \mathrm{d}^2\Omega}{(2\pi)^3} \ \left|\tilde{\delta}(\boldsymbol{k})\right|^2. \tag{3.10}$$

At this stage, we can define the *power spectrum* as the average of $|\tilde{\delta}(\boldsymbol{k})|^2$ over all directions:

$$P(k) \equiv \frac{1}{4\pi} \int \mathrm{d}^2\Omega \, \left|\tilde{\delta}(\boldsymbol{k})\right|^2. \tag{3.11}$$

If we consider only the case in which δ is isotropic, then the power spectrum can be written directly as $P(k) = |\tilde{\delta}(\boldsymbol{k})|^2$. Sometimes, an alternative definition can also be found in the literature:

$$\left\langle \tilde{\delta}(\boldsymbol{k})\tilde{\delta}^*(\boldsymbol{k}') \right\rangle_\Omega \equiv (2\pi)^3 P(k)\delta^{(3)}(\boldsymbol{k} - \boldsymbol{k}'), \tag{3.12}$$

where δ at the right-hand side is the Dirac function. This is strictly equivalent to Eq. (3.11).

Meanwhile, by inserting Eq. (3.11) into Eq. (3.10), we obtain

$$\sigma^2 = \int \frac{4\pi k^2 \mathrm{d}k}{(2\pi)^3} \, P(k) = \int \frac{\mathrm{d}k}{k} \, \Delta^2(k), \tag{3.13}$$

where

$$\Delta^2(k) \equiv \frac{k^3}{2\pi^2} P(k) \tag{3.14}$$

is called the *dimensionless power spectrum*. In most cases, $\Delta^2(k)$ is modelled by power laws. For example, the power spectrum $\Delta_{\mathcal{R}}^2(k)$ of the primordial scalar field $\mathcal{R}$ usually takes the following form:

$$\Delta_{\mathcal{R}}^2(k) = A \left(\frac{k}{k_0} \right)^{n_{\mathrm{s}}-1}, \tag{3.15}$$

where A is the amplitude, n_{s} is the scalar spectral index, and k_0 is a pivot scale arbitrarily chosen. In early years, cosmologists assume that $n_{\mathrm{s}} = 1$, so that the primordial perturbation is scale-independent. The power is identical for all modes. However, recent studies show that n_{s} is slightly smaller than 1. The case $n_{\mathrm{s}} = 1$ has been excluded at 3.6-σ (Hinshaw et al. 2013).

A third definition of the power spectrum can be provided with Eq. (3.13). As mentioned above, the variance σ^2 can actually diverge as Eq. (3.13) extends over $\mathbb{R}$. However, the larger $|k|$ is, the smaller the scale, and the infinitely small scales can not be reached in observations. Therefore, it is convenient to define an upper cutoff for the integral of σ^2. In the end, if we define the cutoff variance $\tilde{\sigma}^2$ as

$$\tilde{\sigma}^2(k) \equiv \int_0^k \frac{\mathrm{d}k'}{k'} \, \Delta^2(k'), \tag{3.16}$$

the dimensionless power spectrum can be considered as the variation of the cutoff variance per logarithmic scale, shown by

$$\Delta^2(k) = \frac{\mathrm{d}\tilde{\sigma}^2(k)}{\mathrm{d}\ln k}. \tag{3.17}$$

3.1.2 Matter fluctuation

Let us focus on the case of the matter field. Let δ be the matter density contrast of the Universe. Instead of setting a upper cutoff, the integral (3.9) can also be regularized by

---

filtering the field. Denote $W_R(\boldsymbol{r})$ as a filter function whose size is characterized by a physical size R. Then, the new quantity

$$\sigma^2(R) \equiv \left\langle (W_R * \delta)^2(\boldsymbol{r}) \right\rangle_{\boldsymbol{r}} \tag{3.18}$$

is finite if W_R has a finite support. Using Eqs. (3.5), (3.10) and (3.11), we find

$$
\begin{aligned}
\sigma^2(R) &= \int \frac{k^2 \mathrm{d}k \, \mathrm{d}^2\Omega}{(2\pi)^3} \left| \left(\widetilde{W_R * \delta} \right)(\boldsymbol{k}) \right|^2 \\
&= \int \frac{k^2 \mathrm{d}k \, \mathrm{d}^2\Omega}{(2\pi)^3} \left| \left(\widetilde{W}_R \cdot \tilde{\delta} \right)(\boldsymbol{k}) \right|^2 \\
&= \int \frac{k^2 \mathrm{d}k}{2\pi^2} \, P(k) \left| \widetilde{W}_R(\boldsymbol{k}) \right|^2,
\end{aligned}
\tag{3.19}
$$

where $P(k)$ is the matter power spectrum. A common choice for W_R is the 3D top-hat filter with radius R: $W_R(\boldsymbol{r}) = (3/4\pi R^3) \cdot \Theta(R - r)$, where Θ is the Heaviside step function. In this case,

$$\widetilde{W}_R(k) = \frac{3}{k^3 R^3} \Big[\sin(kR) - kR\cos(kR) \Big]. \tag{3.20}$$

So, the power spectrum is passed through a window function such that the integration does not diverge anymore. Particularly,

$$\sigma_8^2 \equiv \sigma^2(R = 8 \text{ Mpc}/h) \tag{3.21}$$

represents the variance of the matter density contrast smoothed with a spherical top-hat filter with physical radius 8 Mpc/h. It is a parameter often considered in lensing analyses. Sometimes, the size of the window function is determined by a given "mass scale". That means the radius of the top-hat sphere is set so that the mass of enclosed matter is M. Denote $\bar{\rho}_0$ as the background mass density at the present time. The definition of the mass-based matter fluctuation leads to

$$\sigma^2(M) \equiv \sigma^2 \left(R = \left(\frac{3M}{4\pi\bar{\rho}_0} \right)^{1/3} \right). \tag{3.22}$$

3.2 Linear perturbation theory

The Universe on large enough scales can be considered as homogeneous. However, on smaller scales, primordial fluctuations leave inhomogeneous footprints and get transformed into matter and radiation. When these large scales enter the horizon, local inhomogeneity starts to evolve. In comoving coordinates, the evolution is the combination of two effects: attraction by gravity and repulsion by the fluid pressure. For modes which cross the horizon relatively late, the evolution is still in its early stage and can be well described by linear perturbation theory. This linear regime will be the focus of this section.

Fluid dynamic equations

Consider a universe in the epoch after the matter-radiation equality, so that we can assume a matter- or dark-energy-dominated universe. Following Peebles (1993), we will model the

matter as an ideal pressureless fluid whose state is characterized by the density $\rho(t, \boldsymbol{r})$ and the velocity $\boldsymbol{u}(t, \boldsymbol{r})$, both defined at cosmic time t in physical coordinates $\boldsymbol{r}$. The dynamic of Newtonian physics can de described by

$$\text{Continuity equation:} \quad \frac{\mathrm{d}\rho}{\mathrm{d}t}(t, \boldsymbol{r}) + \nabla \cdot (\rho(t, \boldsymbol{r})\boldsymbol{u}(t, \boldsymbol{r})) = 0, \tag{3.23}$$

$$\text{Euler equation:} \quad \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}(t, \boldsymbol{r}) + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u}(t, \boldsymbol{r}) = -\nabla\Phi(t, \boldsymbol{r}), \tag{3.24}$$

$$\text{Poisson equation:} \quad \Delta\Phi(t, \boldsymbol{r}) = 4\pi\mathrm{G} \sum_\alpha (1 + 3w_\alpha)\rho_\alpha(t, \boldsymbol{r}), \tag{3.25}$$

where Φ is the Newtonian potential, α runs over different cosmological components (matter, photons, neutrinos, dark energy, etc.), and ρ_α and w_α are corresponding density and equation of state of each component.

Equations (3.23) (3.24), and (3.25) represent respectively mass conservation, momentum conservation, and gravity. The right-hand side of Eq. (3.24) has already been reduced to a single gravity term because of considering a pressureless fluid. If we want to establish the dynamic for the radiation-dominated epoch, then the pressure term needs to be accounted for. The unusual form of the Poisson equation is the general expression when all cosmological components are accounted for. Actually, the Poisson equation can be seen as a classical limit of the first Friedmann equation (Eq. 2.5), which is derived from the 00 component of the Einstein equation. If we take $\alpha = \mathrm{m}$ (matter), then $w_\alpha = 0$ and the usual Poisson equation is recovered. If in addition, the cosmological constant is considered, then the right-hand side of Eq. (3.25) becomes $4\pi\mathrm{G}\rho_\mathrm{m} - \mathrm{c}^2\Lambda$, since $w_\mathrm{de} = -1$ and $\rho_\mathrm{de} = \mathrm{c}^2\Lambda/8\pi\mathrm{G}$.

To avoid the possible confusion in the follows, I will use the total derivative with regard to t instead of the partial one. These two notations are equivalent in the Eulerian spatial representation, since a 3D Eulerian point is independent from time.

Change of variables

Let us now write these equations in comoving space, by first introducing some notations. Let $\boldsymbol{R} \equiv \boldsymbol{r}/a$ be the comoving coordinates associated to $\boldsymbol{r}$. Since $\boldsymbol{R}$ actually depends on t via a, the total derivative $\mathrm{d}/\mathrm{d}t$ is not equal to $\partial/\partial t$ anymore. For a scalar field f, denote $f_{\boldsymbol{r}}$ and $f_{\boldsymbol{R}}$ as its mathematical forms in $\boldsymbol{r}$ and $\boldsymbol{R}$ spaces respectively. Then, the change of variables $\boldsymbol{R}(t) = \boldsymbol{r}/a(t)$ leads to

$$f_{\boldsymbol{r}}(t, \boldsymbol{r}) = f_{\boldsymbol{r}}(t, a(t)\boldsymbol{R}(t)) = f_{\boldsymbol{R}}(t, \boldsymbol{R}(t)), \tag{3.26}$$

and

$$\begin{aligned} \frac{\mathrm{d}f_{\boldsymbol{r}}}{\mathrm{d}t}(t, \boldsymbol{r}) &= \frac{\mathrm{d}f_{\boldsymbol{R}}}{\mathrm{d}t}(t, \boldsymbol{R}(t)) \\ &= \frac{\partial f_{\boldsymbol{R}}}{\partial t} + \frac{\partial \boldsymbol{R}}{\partial t} \cdot (\nabla f_{\boldsymbol{R}}) \\ &= \frac{\partial f_{\boldsymbol{R}}}{\partial t} + \frac{\dot{a}}{a^2}\boldsymbol{r} \cdot (\nabla f_{\boldsymbol{R}}) \\ &= \frac{\partial f_{\boldsymbol{R}}}{\partial t} + H(t)\boldsymbol{R}(t) \cdot (\nabla f_{\boldsymbol{R}}), \end{aligned} \tag{3.27}$$

where $\dot{a} = \mathrm{d}a/\mathrm{d}t$ and $H(t) = \dot{a}/a$ is the Hubble parameter. If $\boldsymbol{f}$ is a vector field, then $\boldsymbol{R}(t) \cdot (\nabla f_{\boldsymbol{R}})$ from Eq. (3.27) should be replaced by $(D\boldsymbol{f}_{\boldsymbol{R}})(\boldsymbol{R})$ as the differential of $\boldsymbol{f}_{\boldsymbol{R}}$

applied at $\mathbf{R}$. One verifies at ease that

$$(D\mathbf{f_R})(\mathbf{R}) = (\mathbf{R} \cdot \nabla)\mathbf{f_R} \tag{3.28}$$

by expanding component by component.

On the other hand, the velocity $\mathbf{u}$ describes the total velocity with regard to an Eulerian reference point, which takes also the expansion of the Universe into account. If one extracts the cosmic flow from it, the remaining part is called *peculiar velocity*, denoted here as $\mathbf{v}$. Precisely, one can write

$$\mathbf{u}(t, \mathbf{r}) = \mathbf{v}(t, \mathbf{r}) + H(t) \cdot \mathbf{r}, \tag{3.29}$$

or in comoving coordinates

$$\mathbf{u}(t, \mathbf{R}) = \mathbf{v}(t, \mathbf{R}) + \dot{a}(t) \cdot \mathbf{R}. \tag{3.30}$$

Here, as I will do for the rest of the section, the subscripts $_\mathbf{r}$ and $_\mathbf{R}$ have been omitted. Since in equations we will be dealing with the physical quantities which should be the same in comoving and physical spaces, the confusion does not exist anymore. Particularly, $\mathbf{u}(t, \mathbf{R})$ and $\mathbf{v}(t, \mathbf{R})$ should be interpreted as physical velocities parametrized by comoving coordinates, not comoving velocities.

Finally, the nabla operator here is the gradient derivative which depends implicitly on the space in which the field is parametrized. For example, omitting the subscripts $_\mathbf{r}$ and $_\mathbf{R}$, the change of variables leads to

$$\nabla f(t, \mathbf{r}) = \frac{1}{a(t)} \nabla f(t, \mathbf{R}). \tag{3.31}$$

Dynamic in comoving coordinates

Let us start with writing the continuity equation in comoving space. With Eqs. (3.27) and (3.30), Eq. (3.23) becomes

$$\left(\frac{\partial \rho}{\mathrm{d}t}(t, \mathbf{R}) - H\mathbf{R} \cdot \nabla\rho(t, \mathbf{R})\right) + \frac{1}{a}\nabla \cdot \left(\rho(t, \mathbf{R})(\mathbf{v}(t, \mathbf{R}) + \dot{a}\mathbf{R})\right) = 0. \tag{3.32}$$

After simplification, we obtain

$$\frac{\partial \rho}{\partial t} - 3H\rho + \frac{1}{a}\nabla \cdot (\rho\mathbf{v}) = 0, \tag{3.33}$$

where $\nabla \cdot \mathbf{R} = 3$ is needed.

Then, we derive an Euler equation for $\mathbf{v}(t, \mathbf{R})$. Applying Eqs. (3.27) and (3.30), we obtain

$$\frac{\partial \mathbf{u}}{\partial t}(t, \mathbf{R}) - \frac{\dot{a}}{a}(D\mathbf{u}) \cdot \mathbf{R} + \frac{1}{a}\Big((\mathbf{v} + \dot{a}\mathbf{R}) \cdot \nabla\Big)\mathbf{u}(t, \mathbf{R}) = -\frac{1}{a}\nabla\Phi(t, \mathbf{R}), \tag{3.34}$$

and the left-hand side can be simplified by Eq. (3.28), leading to

$$\frac{\partial \mathbf{u}}{\partial t}(t, \mathbf{R}) + \frac{1}{a}(\mathbf{v} \cdot \nabla)\mathbf{u}(t, \mathbf{R}) = -\frac{1}{a}\nabla\Phi(t, \mathbf{R}). \tag{3.35}$$

The first term from Eq. (3.35) involves $\partial(\dot{a}\boldsymbol{R})/\partial t$. Although $\boldsymbol{R}$ depends on t, its dependence should be neglected in the partial derivative. Therefore, $\partial(\dot{a}\boldsymbol{R})/\partial t = \ddot{a}\boldsymbol{R}$. Using Eq. (3.30) and $(\boldsymbol{v} \cdot \nabla)\boldsymbol{R} = \boldsymbol{v}$, we get

$$\frac{\partial \boldsymbol{v}}{\partial t} + \ddot{a}\boldsymbol{R} + H\boldsymbol{v} + \frac{1}{a}(\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\frac{1}{a}\nabla\Phi. \tag{3.36}$$

For the Poisson equation, it is useful to decompose the potential into homogeneous and inhomogeneous parts, such that

$$\Delta\Phi(t, \boldsymbol{r}) = \Delta\phi(t, \boldsymbol{r}) + 4\pi G \sum_\alpha (1 + 3w_\alpha)\bar{\rho}_\alpha(t) \tag{3.37}$$

with the *reduced Newtonian potential* ϕ:

$$\Delta\phi(t, \boldsymbol{r}) \equiv 4\pi G(\rho(t, \boldsymbol{r}) - \bar{\rho}(t)), \tag{3.38}$$

where $\rho = \rho_{\mathrm{m}}$ for simplicity. Moreover, $\rho - \bar{\rho} = \bar{\rho}\delta = \bar{\rho}_0 a^{-3}\delta$ and $\Omega_{\mathrm{m}} = 8\pi G\bar{\rho}_0/3H_0^2$, so $4\pi G(\rho - \bar{\rho}) = 4\pi G\bar{\rho}_0\delta a^{-3} = 3H_0^2\Omega_{\mathrm{m}}\delta/2a^3$. Thus, Eq. (3.38) becomes

$$\Delta\phi(\boldsymbol{r}) = \frac{3H_0^2\Omega_{\mathrm{m}}}{2}\frac{\delta}{a^3}. \tag{3.39}$$

The second term at the right-hand side of Eq. (3.37) is independent from $\boldsymbol{r}$ since all components except for matter are assumed to be distributed uniformly. This homogeneity can be described by the Friedmann equation (2.6). In comoving coordinates, we end up with

$$\frac{1}{a^2}\Delta\Phi(t, \boldsymbol{R}) = \frac{3H_0^2\Omega_{\mathrm{m}}}{2a^3}\delta(t, \boldsymbol{R}) - \frac{3\ddot{a}}{a}. \tag{3.40}$$

It is convenient to replace $\rho(t, \boldsymbol{R})$ by $\bar{\rho}(t)(1 + \delta(t, \boldsymbol{R}))$ and to reason in the density contrast. This can simplify Eq. (3.33) knowing that $\bar{\rho}(t) = \bar{\rho}_0 a^{-3}(t)$. At the end of the day, the dynamic equations for δ and $\boldsymbol{v}$ parametrized by comoving coordinates are

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot ((1 + \delta)\boldsymbol{v}) = 0, \tag{3.41}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + \ddot{a}\boldsymbol{R} + H\boldsymbol{v} + \frac{1}{a}(\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\frac{1}{a}\nabla\Phi, \tag{3.42}$$

$$\frac{1}{a^2}\Delta\Phi = \frac{3H_0^2\Omega_{\mathrm{m}}}{2a^3}\delta - \frac{3\ddot{a}}{a}. \tag{3.43}$$

Linearization

The system of equations above describes the evolution of the density contrast with time. However, it is nonlinear and the analytical solution is hard to construct. Using perturbation theory, we can work on the linearized system. The obtained solution will be a good approximation when δ and $\boldsymbol{v}$ stay small. This corresponds to the early stage of the evolution, therefore to large scales which enter the horizon late. In other words, if we cut the Universe into some large enough pieces and associate a density contrast to each of them, then the interaction between these regions will be triggered late, and the contrast field will be well described by the linearized equations showed as follows.

Dropping terms with $\delta \boldsymbol{v}$ and $\boldsymbol{v}^2$, Eqs. (3.41), (3.42), and (3.43) become

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot \boldsymbol{v} = 0, \tag{3.44}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + \ddot{a}\boldsymbol{R} + H\boldsymbol{v} = -\frac{1}{a}\nabla\Phi, \tag{3.45}$$

$$\frac{1}{a^2}\Delta\Phi = \frac{3H_0^2\Omega_{\mathrm{m}}}{2a^3}\delta - \frac{3\ddot{a}}{a}. \tag{3.46}$$

One can deduce easily

$$\begin{aligned}
\frac{\partial}{\partial t}\left(a\frac{\partial\delta}{\partial t}\right) &= -\nabla \cdot \frac{\partial \boldsymbol{v}}{\partial t} \\
&= \ddot{a}\nabla \cdot \boldsymbol{R} + H\nabla \cdot \boldsymbol{v} + \frac{1}{a}\Delta\Phi \\
&= 3\ddot{a} - \dot{a}\frac{\partial\delta}{\partial t} + \frac{3H_0^2\Omega_{\mathrm{m}}}{2a^2}\delta - 3\ddot{a},
\end{aligned} \tag{3.47}$$

which results in

$$\frac{\partial^2\delta}{\partial t^2} + 2H\frac{\partial\delta}{\partial t} - \frac{3H_0^2\Omega_{\mathrm{m}}}{2a^3}\delta = 0. \tag{3.48}$$

This is the linearized equation for the mass density contrast, an ordinary differential equation which admits the general solution of the form (Peebles 1993; Peacock 1999; Dodelson 2003)

$$\delta(t, \boldsymbol{R}) = D_+(t)\delta_+(\boldsymbol{R}) + D_-(t)\delta_-(\boldsymbol{R}), \tag{3.49}$$

where δ_+ and δ_- are two independent particular solutions, which should be interpreted as proportional to the contrast field of the considered scale at horizon crossing. The coefficients of Eq. (3.47) indicates that it exists necessarily a *growing mode*, labelled with +, and a *decaying mode*, labelled with -.

In the linear regime, the density contrast will keep track on the initial condition, only evolving proportionally with the time-dependent factors D_+ and D_-. Supposing that the decaying mode vanishes fast enough for relevant t, the density contrast can then be simply expressed by the growing mode whose general form is

$$D_+(t) - D_+(t_{\mathrm{eq}}) \propto H(t)\int_{t_{\mathrm{eq}}}^t \frac{\mathrm{d}t}{\dot{a}^2(t)}, \tag{3.50}$$

where t_{eq} is the time of the matter-radiation equality. D_+ is called the *growth factor*, and is conventionally normalized so that for a flat universe, $D_+(a) = a$ during the matter-dominated era. In this case, we can write, for $a > a_{\mathrm{eq}}$,

$$D_+(a) = \frac{5\Omega_{\mathrm{m}}}{2}\frac{H(a)}{H_0}\int_0^a \frac{\mathrm{d}a'}{(a'H(a')/H_0)^3}. \tag{3.51}$$

Several examples of D_+ evolution are shown in Fig. 3.1. It is also useful to normalize the growth factor with regard to its value at the current time. Define

$$D(z) = \frac{D_+(a = (1+z)^{-1})}{D_+(a = 1)}, \tag{3.52}$$

then $D(z = 0) = 1$ and the linear density contrast at redshift z can be simply written as $\delta(z, \boldsymbol{R}) = D(z)\delta(\boldsymbol{R})$, where $\delta(\boldsymbol{R})$ represents the fluctuation at the present time.
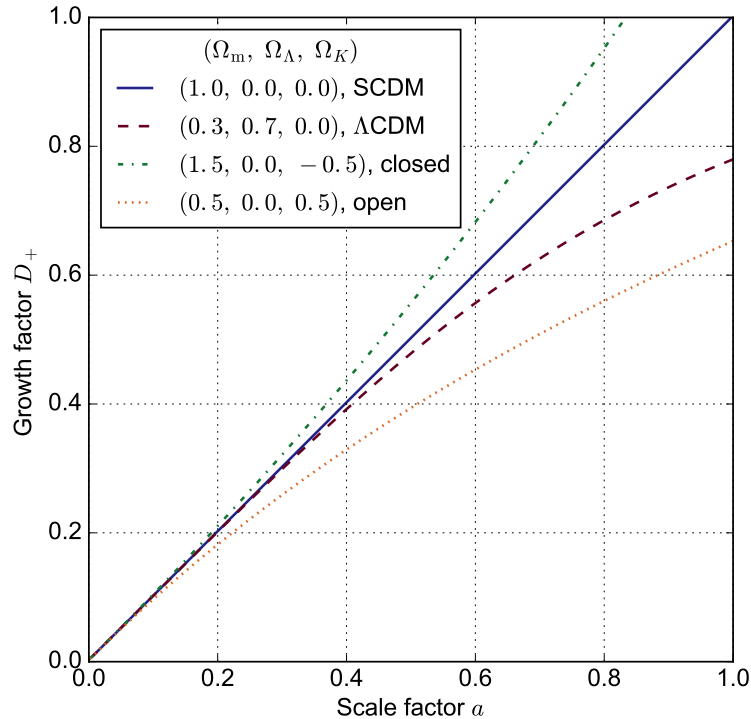
FIGURE 3.1: Evolution of the growth factor under different cosmologies. In the standard cold dark matter (SCDM) model, we have $D_+(a) = a$.

3.3 Spherical collapse model

Concerning the nonlinear evolution of the Universe, cosmologists observe from N-body simulations (e.g. Hahn et al. 2007; Chen et al. 2015) that initial perturbations form sheet-like structures by self-gravitation. Then, sheets fold into filaments, and filaments collapse to form bound objects at the end. In ΛCDM-cosmology, these bound objects are composed principally of dark matter, called *dark-matter halos*. The halos contain also baryons which leads to forming galaxies. Since dark matter is not visible, what we observe is clusters of galaxies. For this reason, clusters (of matter) and halos are often interchangeable in the literature.

The formation of the halos in N-body simulations is complex. Here, we take a simplified approach. We will skip sheets and filaments, and use the spherical collapse model to explain the cluster formation. Following Peebles (1980), we consider a globally homogeneous and isotropic universe and a spherical region inside. We take out the mass contained within a thin shell of the sphere, and add it to the inner region, so that the inner mass density is slightly higher than the background. We further assume that the background universe is flat and matter-dominated, so that $\Omega_{\mathrm{m}}^{(\mathrm{b})} = 1$ and $\Omega_K^{(\mathrm{b})} = 0$. The spherical region is overdense and can be modelled as a closed universe with $0 < -\Omega_K = \Omega_{\mathrm{m}} - 1 \ll 1$ (Fig. 3.2).

We are going to solve the Friedmann equation to obtain the evolution of the scale factor. Contrary to the prescription from Chap. 2 where all densities have been scaled and expressed with the fraction at the present time, here we will scale to the initial time, since $\Omega_{\mathrm{m}}, \Omega_K, \Omega_{\mathrm{m}}^{(\mathrm{b})}$, and $\Omega_K^{(\mathrm{b})}$ are defined at the moment where the "split" takes place. Thus, in this section, t_0 is refered to the initial time, with $a(t_0) = 1$ and $H_0 = \dot{a}(t_0)/a(t_0)$ to simplify the calculation.
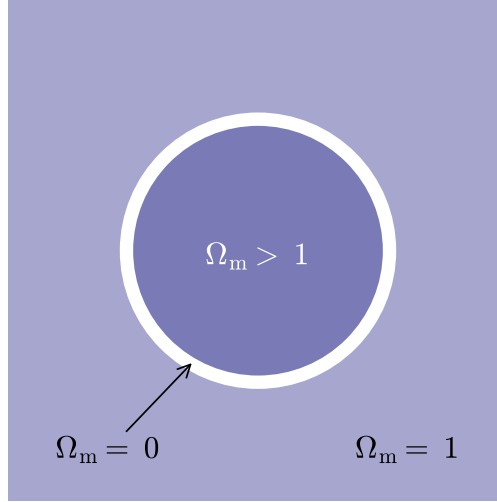
FIGURE 3.2: Illustration of an idealized scenario for the spherical collapse. In an homogeneous Einstein-de Sitter universe ($\Omega_\mathrm{m} = 1$), we empty the mass on a thin shell and put it inside, so that $\Omega_\mathrm{m} > 1$ for the inner region.

Collapse: evolution of a closed universe

For the overdense region, Friedmann's equation (Eq. 2.19) leads to

$$\left(\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 = H_0^2 \left(\frac{\Omega_\mathrm{m}}{a} + (1 - \Omega_\mathrm{m})\right). \tag{3.53}$$

Let $\theta = H_0 \tau \sqrt{\Omega_\mathrm{m} - 1}$ where τ is the conformal time in the overdense region. The derivative of a with regard to θ is

$$\left(\frac{\mathrm{d}a}{\mathrm{d}\theta}\right)^2 = \left(\frac{\mathrm{d}t}{\mathrm{d}\theta}\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 = \frac{a^2}{H_0^2(\Omega_\mathrm{m} - 1)} \cdot H_0^2 \left(\frac{\Omega_\mathrm{m}}{a} + (1 - \Omega_\mathrm{m})\right)$$

$$= \frac{\Omega_\mathrm{m} a}{\Omega_\mathrm{m} - 1} - a^2. \tag{3.54}$$

To obtain the evolution of a with regard to θ, we need to rearrange Eq. (3.54) and integrate. Denote $A = \Omega_\mathrm{m}/2(\Omega_\mathrm{m} - 1)$ and $b = a/A - 1$, the change of variables leads to

$$\mathrm{d}\theta = \frac{\mathrm{d}a}{\sqrt{-(a - A)^2 + A^2}} = \frac{1}{A}\frac{\mathrm{d}a}{\sqrt{1 - (a/A - 1)^2}} = \frac{\mathrm{d}b}{\sqrt{1 - b^2}} = -\mathrm{d}(\arccos b). \tag{3.55}$$

The primitive function of $1/\sqrt{1 - b^2}$ can be $\arcsin b$ or $\arccos b$. The latter has been chosen because we wish θ to be positive. Therefore, the scale factor of the overdense region parametrized by θ is

$$a(\theta) = \frac{\Omega_\mathrm{m}}{2(\Omega_\mathrm{m} - 1)}(1 - \cos\theta) \tag{3.56}$$

for $\theta_0 \leq \theta < 2\pi$, where θ_0 satisfies $a(\theta_0) = 1$. Then, from the relation between θ and τ, we can derive $a\mathrm{d}\theta = H_0\sqrt{\Omega_\mathrm{m} - 1}\,\mathrm{d}t$, so

$$t(\theta) = \frac{\Omega_\mathrm{m}}{2H_0(\Omega_\mathrm{m} - 1)^{3/2}}(\theta - \sin\theta) \tag{3.57}$$

for $\theta_0 \leq \theta < 2\pi$. Equation (3.56) shows that the size of the overdense region increases first, reaches a maximum at $\theta = \theta_{\max} = \pi$, and decreases to 0 at the end. This means that the closed universe will expand until a turn-around point, then collapse. At the turn-around point, the corresponding corresponding scale factor and time are

$$a_{\max} = \frac{\Omega_{\mathrm{m}}}{\Omega_{\mathrm{m}} - 1}, \tag{3.58}$$

$$t_{\max} = \frac{\pi}{2H_0} \frac{\Omega_{\mathrm{m}}}{(\Omega_{\mathrm{m}} - 1)^{3/2}}. \tag{3.59}$$

We will use these two quantities to normalize a and t, which yield the normalized scale factor $\tilde{a}$ and the dimensionless time $\tilde{t}$:

$$\tilde{a}(\theta) \equiv \frac{a(\theta)}{a_{\max}} = \frac{1 - \cos\theta}{2}, \tag{3.60}$$

$$\tilde{t}(\theta) \equiv \frac{t(\theta)}{t_{\max}} = \frac{\theta - \sin\theta}{\pi}. \tag{3.61}$$

Meanwhile, the background evolves as a matter-dominated flat universe. The solution for the Friedmann equation is

$$a^{(\mathrm{b})}(t) = \left(\frac{3H_0 t}{2}\right)^{2/3}, \tag{3.62}$$

where $t \geq t_0$. After normalization, for $\theta > \theta_0$, the background scale factor is

$$\begin{aligned}
\tilde{a}^{(\mathrm{b})}(\theta) \equiv \frac{a^{(\mathrm{b})}(\theta)}{a_{\max}} &= \frac{\Omega_{\mathrm{m}} - 1}{\Omega_{\mathrm{m}}} \left(\frac{3H_0 t(\theta)}{2}\right)^{2/3} \\
&= \frac{\Omega_{\mathrm{m}} - 1}{\Omega_{\mathrm{m}}} \left(\frac{3H_0}{2} \frac{t(\theta)}{t_{\max}} \cdot \frac{\pi}{2H_0} \frac{\Omega_{\mathrm{m}}}{(\Omega_{\mathrm{m}} - 1)^{3/2}}\right)^{2/3} \\
&= \frac{1}{\Omega_{\mathrm{m}}^{1/3}} \left(\frac{3\pi}{4} \tilde{t}(\theta)\right)^{2/3} \approx \left(\frac{3}{4}(\theta - \sin\theta)\right)^{2/3},
\end{aligned} \tag{3.63}$$

since $\Omega_{\mathrm{m}} \approx 1$.

Mass overdensity

Now, we would like to know the evolution of the *mass overdensity* Δ and the density contrast δ described by the spherical collapse model. These two quantities are defined as

$$\Delta \equiv 1 + \delta = \frac{\rho(a)}{\rho^{(\mathrm{b})}(a^{(\mathrm{b})})} = \frac{\rho_0 a^{-3}}{\rho_0^{(\mathrm{b})} a^{(\mathrm{b})-3}}. \tag{3.64}$$

The ratio of mass density at the initial time t_0 is nothing but $\Omega_{\mathrm{m}}/\Omega_{\mathrm{m}}^{(\mathrm{b})}$, which is close to 1. As a result, with Eqs. (3.60) and (3.63),

$$\Delta = 1 + \delta \approx \left(\frac{\tilde{a}^{(\mathrm{b})}}{\tilde{a}}\right)^3 = \frac{9(\theta - \sin\theta)^2}{2(1 - \cos\theta)^3}. \tag{3.65}$$

What is the density contrast at $t_{\max}$? If we take $\theta = \theta_{\max} = \pi$, then

$$1 + \delta_{\max} = \frac{9\pi^2}{16} \approx 5.55, \tag{3.66}$$

which means that the collapse starts when the density contrast reaches about 5.55.

We can observe that in this model, if $a \to 0$, then $\delta \to \infty$. However, in the real universe, the collapse will stop before an infinite density is reached. One of the ways to regularize this problem is to assume the final state predicted by the virial theorem for the system. This means that the kinetic energy E_k and the potential energy E_p will satisfy

$$E_k = -\frac{1}{2}E_p. \tag{3.67}$$

It is logical to suppose that $E_k = 0$ at the turn-around point because when collapse starts, the "speed" $\dot{a}$ is zero. Let $r_{\max}$ be the physical radius of the region at $t_{\max}$. The potential energy, which is also the total energy, leads to

$$E = E_p = -\frac{3GM^2}{5r_{\max}}. \tag{3.68}$$

It is not difficult to find that at $r_{\mathrm{vir}} = r_{\max}/2$, $E - E_p$ and E_p verifies Eq. (3.67). This condition imposes that the collapse stops at the half of its maximal size. Let us set a_{vir} to $a_{\max}/2$. The corresponding θ_{vir} turns out to be $3\pi/2$. Thus, Eq. (3.65) yields

$$1 + \delta_{\mathrm{vir}} = \frac{9}{2}\left(\frac{3\pi}{2} + 1\right)^2 \approx 147. \tag{3.69}$$

However, for historical reasons, cosmologists derive the virial overdensity differently. Let $\theta_{\mathrm{end}} = 2\pi$ be the value of the time parameter at the end of the evolution. By taking $\theta = \theta_{\mathrm{end}}$ in the numerator of Eq. (3.65) and $\theta = \theta_{\mathrm{vir}}$ in the denominator, this strange choice leads to

$$1 + \delta_{\mathrm{end}} \equiv \frac{9(\theta_{\mathrm{end}} - \sin\theta_{\mathrm{end}})^2}{2(1 - \cos\theta_{\mathrm{vir}})^3} = 18\pi^2 \approx 178, \tag{3.70}$$

which is the result more commonly found in the literature. Whatever the numerical result is, the conclusion is the following: perturbations form gravitationally bound structures when they become about 147–178 times denser than the background. This simple model is consistent with N-body simulations.

Collapse in linear theory

At the end of this section, we will analyze the collapse in linear theory. What would be the linear part of the evolution when the spherical collapse is about to start? To answer to this question, let us develop a and t in power series of θ. The results are

$$\tilde{a}_{\mathrm{lin}} \equiv \frac{a_{\mathrm{lin}}}{a_{\max}} = \frac{1}{4}\theta^2 - \frac{1}{48}\theta^4, \tag{3.71}$$

$$\tilde{t}_{\mathrm{lin}} \equiv \frac{t_{\mathrm{lin}}}{t_{\max}} = \frac{1}{6\pi}\theta^3 - \frac{1}{120\pi}\theta^5. \tag{3.72}$$

Let us try to express $\tilde{a}_{\text{lin}}$ in terms of $\tilde{t}_{\text{lin}}$. Denote $x = 6\pi\tilde{t}_{\text{lin}}$. Applying a linear approximation to Eq. (3.72):

$$\theta^3 = x + \frac{1}{20}\theta^5 \approx x + \frac{1}{20}x^{5/3} = x\left(1 + \frac{1}{20}x^{2/3}\right), \tag{3.73}$$

and inserting it into Eq. (3.71), we find

$$\begin{aligned}
\tilde{a}_{\text{lin}} &= \frac{1}{4}\theta^2\left(1 - \frac{1}{12}\theta^2\right) \\
&\approx \frac{1}{4}x^{2/3}\left(1 + \frac{1}{20}x^{2/3}\right)^{2/3}\left(1 - \frac{1}{12}x^{2/3}\left(1 + \frac{1}{20}x^{2/3}\right)^{2/3}\right) \\
&\approx \frac{1}{4}x^{2/3}\left(1 + \frac{1}{30}x^{2/3}\right)\left(1 - \frac{1}{12}x^{2/3}\right) \\
&\approx \frac{1}{4}x^{2/3}\left(1 - \frac{1}{20}x^{2/3}\right).
\end{aligned} \tag{3.74}$$

Therefore, the density contrast in linear theory is

$$\begin{aligned}
1 + \delta_{\text{lin}} = \left(\frac{\tilde{a}^{(b)}}{\tilde{a}_{\text{lin}}}\right)^3 &= \left(\frac{1}{4}x^{2/3}\right)^3\left(\frac{1}{4}x^{2/3}\left(1 - \frac{1}{20}x^{2/3}\right)\right)^{-3} \\
&\approx 1 + \frac{3}{20}x^{2/3} \\
&= 1 + \frac{3}{20}\left(6\pi\tilde{t}(\theta)\right)^{2/3}.
\end{aligned} \tag{3.75}$$

Taking $\theta = \theta_{\text{max}} = \pi$, we see that the linear density contrast at the turn-around point is $1 + \delta_{\text{max,lin}} \approx 2.06$. Taking $\theta = \theta_{\text{end}} = 2\pi$, the linear density contrast at the end of the collapse is $1 + \delta_{\text{end,lin}} = 1 + (3/20)(12\pi)^{2/3} \approx 2.686$. This value $\delta_{\text{c}} \equiv 1.686$ is usually considered as a *density contrast threshold* above which a sub-region of the universe is assumed collapsed, with the actual overdensity $\Delta = 178$.

3.4 Mass function

3.4.1 The Press-Schechter formalism

The abundance of the halos is characterized by the *mass function*, which depends on mass and redshift. It is an important halo statistic for cosmology. A simple way to model the mass function is proposed by Press & Schechter (1974, see also Peacock 1999 for interpretation). Consider the density contrast threshold δ_{c} in the linear theory and a contrast field δ smoothed with a top-hat filter of size r. The reasoning of Press & Schechter (1974) is that a (Lagrangian) point is supposed to fall into a collapsed region if the smoothed contrast is larger than $\delta_{\text{c}}/\sigma(r)$, and the characteristic size of each of these regions is larger than r. More precisely, given a mass threshold M, and the fluctuation level $\sigma(M)$ defined by this mass (Eq. 3.22), the total mass represented by a smoothed contrast larger than $\delta_{\text{c}}/\sigma(M)$ would be organized as a cluster with mass larger than M.

However, there exists a small problem with this approach: the underdense regions are not taken into account. This can be easily shown by taking $M \to 0$. When M is arbitrarily small, the total mass above M is just the mass of the Universe, but since $\sigma(M)$ becomes

arbitrarily large, the mass represented by Lagrangian points with smoothed contrast larger than $\delta_c/\sigma(M)$ only accounts for a part of the mass in the Universe. To fix this discrepancy, Press & Schechter (1974) simply propose to add a factor of 2, since they model the distribution of δ of Gaussian random field such that the positive and the negative part are equal. As a result, by denoting $n(z, M)\mathrm{d}M$ as the halo number density for mass within $[M, M+\mathrm{d}M[$, we have

$$\int_M^{+\infty} \mathrm{d}M \; \frac{Mn(z,M)}{\bar{\rho}(z)} = 2 \int_\nu^{+\infty} \frac{\mathrm{d}u}{\sqrt{2\pi}} \; \exp\left(-\frac{u^2}{2}\right), \tag{3.76}$$

where $\nu \equiv \delta_c/D(z)\sigma(M)$ which takes into account the linear redshift dependence, $\bar{\rho}(z)$ is the mean density at z, and u should be understood as the smoothed density contrast [♯1]. Deriving both sides with regard to $\ln\nu$, we obtain

$$\frac{Mn(z,M)}{\bar{\rho}(z)} \frac{\mathrm{d}M}{\mathrm{d}\ln\nu} = \sqrt{\frac{2}{\pi}} \; \nu\exp\left(-\frac{\nu^2}{2}\right). \tag{3.77}$$

In the literature, there exist different conventions for noting a mass function. In this thesis, I will adopt the one used by Jenkins et al. (2001). First, let $n(z, < M)$ be the number of halos with mass smaller than M. Immediately, we have $\mathrm{d}n(z, < M) = n(z, M)\mathrm{d}M$. Now, let us define the *mass function* as

$$f(\nu) \equiv \frac{M}{\bar{\rho}(z)} \frac{\mathrm{d}n(z, < M)}{\mathrm{d}\ln\nu}. \tag{3.78}$$

The interpretation of f is straightforward with Eq. (3.76): it characterizes the correspondence between virialized objects of mass M and linear contrasts in log scale, and the issue of negative contrasts leads to impose the condition $\int \mathrm{d}(\ln\nu) \; f(\nu) = 1$ for f. Sometimes f is called the *multiplicity function*. With the prescription of Eq. (3.78), the Press-Schechter mass function is

$$f_{\mathrm{PS}}(\nu) = \sqrt{\frac{2}{\pi}} \; \nu\exp\left(-\frac{\nu^2}{2}\right). \tag{3.79}$$

In the end, the halo number density depends on cosmology via the Friedmann equation and the power spectrum. The former governs the evolution of $\bar{\rho}(z)$, and the latter is essential for determining the level of $\sigma(M)$.

3.4.2 Other models

Since then, better mass function models, fitted by N-body simulations, have been proposed. Sheth & Tormen (1999, 2002) gave the following expression which stands for the ellipsoidal collapse model:

$$f_{\mathrm{ST}}(\nu) = A \sqrt{\frac{a\nu^2}{2\pi}} \cdot \left(1 + \left(a\nu^2\right)^{-p}\right) \exp\left(-\frac{a\nu^2}{2}\right), \tag{3.80}$$

[♯1]The interpretations of both sides of Eq. (3.76) are not physically consistent. While the left-hand side stands for a mass fraction, the right-hand side is a fraction of volume. A more rigorous approach can be obtained by adding a factor of $1 + \sigma(M) \cdot u$ to the integrand of the right-hand side

where $p = 0.3$, $a = 0.75$ and A is the normalization constant, which is about 0.32218 in this case. If we set $p = 0$ and $a = 1$, A turns out to be 2, and the case of Press & Schechter (1974) is recovered.

Jenkins et al. (2001) provided a fitted model by using the Virgo Consortium simulations, leading to

$$f_{\mathrm{J01}}(\nu = \delta_{\mathrm{c}}/x) = 0.315 \cdot \exp\left(-|\ln x^{-1} + 0.61|^{3.8}\right), \tag{3.81}$$

where $x = \delta_{\mathrm{c}}/\nu = D(z)\sigma(M)$.

Warren et al. (2006) derived their model from the following fitting form:

$$f_{\mathrm{W06}}(\nu) = 0.7234 \cdot \left(x^{-1.625} + 0.2538\right) \exp\left(-\frac{1.1982}{x^2}\right). \tag{3.82}$$

Tinker et al. (2008) found that the universality of the mass function is broken, providing a redshift-dependent model:

$$f_{\mathrm{T08}}(\nu) = A(z)\left(\left(\frac{x}{b(z)}\right)^{-a(z)} + 1\right)\exp\left(-\frac{c}{x^2}\right), \tag{3.83}$$

where $A(z) = 0.186/(1+z)^{0.14}$, $a(z) = 1.47/(1+z)^{0.06}$, $b(z) = 2.57 \cdot (1+z)^{\alpha(200)}$, $\log \alpha(\Delta) = -(0.75/\log(\Delta/75))^{1.2}$, and $c = 1.19$.

Bhattacharya et al. (2011) generalized the model from Sheth & Tormen (1999) with redshift dependency:

$$f_{\mathrm{B11}}(\nu) = A(z)\sqrt{\frac{2}{\pi}}\left(a(z)\nu^2\right)^{q/2}\left(1 + \left(a(z)\nu^2\right)^{-p}\right)\exp\left(-\frac{a(z)\nu^2}{2}\right), \tag{3.84}$$

with $A(z) = 0.333/(1+z)^{0.11}$, $a(z) = 0.788/(1+z)^{0.01}$, $p = 0.807$, and $q = 1.795$.

Figure 3.3 shows some mass function models. I would like to highlight that the y-axis, which is the halo number density at redshift z, is determined as

$$n(z, \log M) \equiv \frac{\mathrm{d}n(z, <M)}{\mathrm{d}\log M} = \ln 10 \cdot \frac{\bar{\rho}(z)}{M}\frac{\mathrm{d}\ln \nu}{\mathrm{d}\ln M}f(\nu). \tag{3.85}$$

This is the volume density within a log mass width $\mathrm{d}\log M$. Thus, the unit of $n(z, \log M)$ is $(\mathrm{Mpc}/h)^{-3}$ as physical length because of $\bar{\rho}(z)$, while the unit of $n(z, M)$ is actually $h^4/(M_\odot \mathrm{Mpc}^3)$. The other subtlety is the factor $\ln 10 \approx 2.3$, which is indispensable if the x-axis is displayed in decimal logarithmic scale.

3.5 Halo density profile

3.5.1 The NFW profiles

Let us address the inner structure of a gravitationally bound object. When structures collapse, the distribution of the matter can depend on the position with regard to the center, thus yields a density profile. Using N-body simulations, Navarro et al. (1995, 1996, 1997) provided a halo density profile which is given as follows,

$$\rho_{\mathrm{NFW}}(\boldsymbol{r}) = \frac{\rho_{\mathrm{s}}}{(r/r_{\mathrm{s}})(1 + r/r_{\mathrm{s}})^2}. \tag{3.86}$$
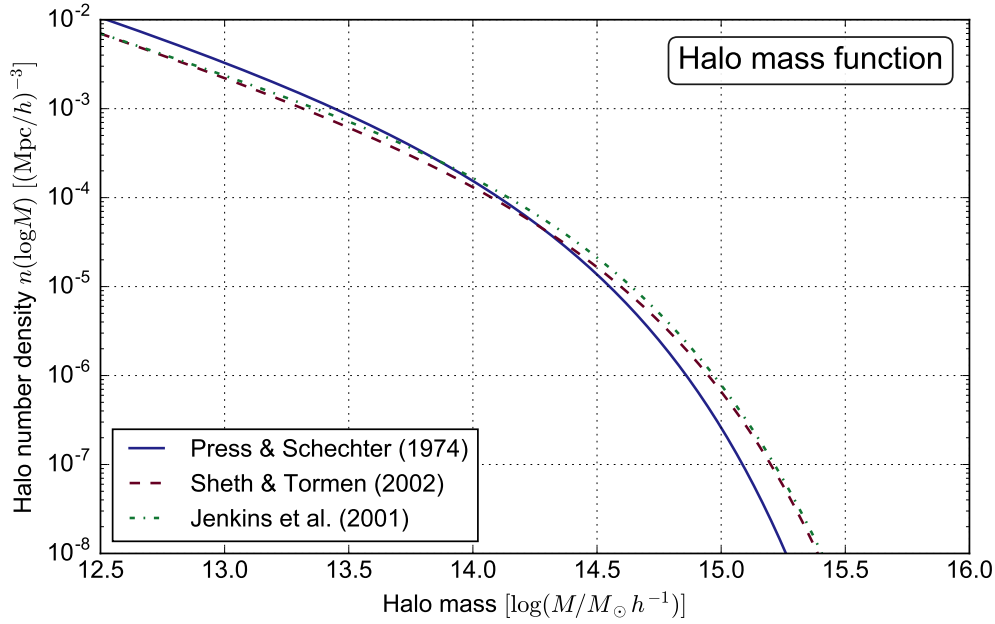
FIGURE 3.3: Various mass function models.

The Navarro-Frenk-White (NFW) profile is parametrized by two numbers: the central mass density ρ_s and the scalar radius r_s. Depending on the convention, these two quantities can have different definitions. A universal way to express them is as follows:

$$\rho_s \equiv \rho_{\rm ref}\Delta \cdot \frac{fc^3}{3}, \tag{3.87}$$

$$r_s \equiv \frac{r_\Delta}{c}. \tag{3.88}$$

The quantity $\rho_{\rm ref}\Delta$ is the mass density inside the halo region, which is Δ times larger than the reference density $\rho_{\rm ref}$. The corresponding radius for the considered Δ is denoted as r_Δ. The ρ_s and r_s are related by the concentration parameter c and a derived quantity f given by

$$f \equiv \frac{1}{\ln(1+c) - c/(1+c)}. \tag{3.89}$$

Choice of the convention

Depending on choices, $\rho_{\rm ref}$ may be $\rho_{\rm crit,0}$, $\rho_{\rm crit}(z)$, $\bar{\rho}_{\rm m,0}$, or $\bar{\rho}_{\rm m}(z)$ (critical and matter densities), and Δ may be a redshift-dependent relation $\Delta_{\rm vir}(z)$ or a constant such as 200 or 500. There is no agreement or consensus on this topic. For example, studies on the mass function tend to use $\bar{\rho}_{\rm m}(z)$, because dark energy, which becomes dominant in the late-time Universe, is not involved in gravitational collapse. However, studies on the mass-concentration relation (Duffy et al. 2008; De Boni et al. 2013; Bhattacharya et al. 2013; Meneghetti et al. 2014; Umetsu et al. 2014; Merten et al. 2015; Diemer & Kravtsov 2014, 2015) prefer to use $\rho_{\rm crit}(z)$ as the reference density. One reason for this less physical choice is historical, since during long time $\Omega_{\rm de}$ had been considered close to 0. The other reason is that $\rho_{\rm crit}$ is higher than $\bar{\rho}_{\rm m}$, so the associated radius is smaller (e.g. $r_{\rm 200c} < r_{\rm 200m}$) and smaller radii lead to a better definition of halo mass (less unbound particles at the outskirt) and scale radius. Particularly, Diemer
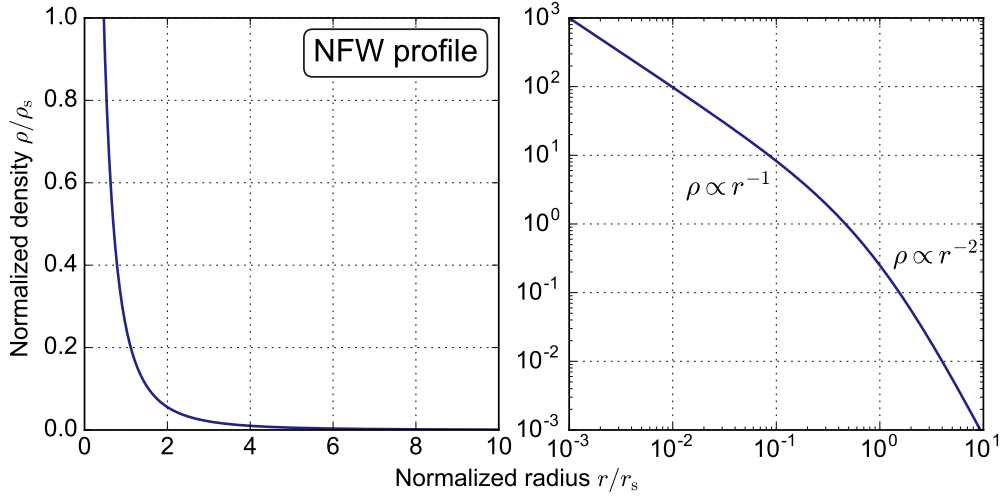
FIGURE 3.4: NFW profile in normalized coordinates. The concentration is set to 10 as an example. The left panel shows the profile in real space, and the right panel represents the same plot in log space.

& Kravtsov (2014) pointed out that the universality of the NFW profiles is better when the density and the radius are scaling respectively by $\rho_{\rm crit}$ and $r_{\rm 200c}$.

Concerning the choice of Δ, a common one is 200 as an approximation of $18\pi^2$ obtained from Eq. (3.70). Many X-ray studies usually pick up $\Delta = 500$, for the same reason as having $\rho_{\rm crit}$: enhance the threshold and reduce the radius to focus on the inner region. However, we can also adopt a cosmology-dependent virial overdensity (see e.g. Bryan & Norman 1998; Henry 2000; Weinberg & Kamionkowski 2003). This is usually fitted from calculations of collapses under different cosmologies.

In this thesis, I adopt the following choice for the NFW profiles:

$$r_{\rm s} = \frac{r_{\rm vir}}{c}, \quad \rho_{\rm s} = \bar{\rho}_{\rm m}(z)\Delta_{\rm vir}(z)\cdot\frac{fc^3}{3} \quad \text{and} \quad \Delta_{\rm vir}(z) \equiv \frac{\rho_{\rm vir}(z)}{\bar{\rho}_{\rm m}(z)}. \tag{3.90}$$

By the definition of the virial density, the mass of a halo is nothing but

$$M = \bar{\rho}_{\rm m}(z)\Delta_{\rm vir}(z)\cdot\frac{4}{3}\pi r_{\rm vir}^3 = 4\pi \int_0^{r_{\rm vir}} r^2 {\rm d}r\ \rho_{\rm NFW}(r). \tag{3.91}$$

The second equality also explains the factor of $fc^3/3$ from Eq. (3.87). Note that the mass is actually not defined if the integral extends to infinity. For this reason, it is useful to define the truncated NFW profiles (labelled as `TJ` for Takada & Jain 2003b):

$$\rho_{\rm TJ}(\boldsymbol{r}) = \frac{\rho_{\rm s}}{(r/r_{\rm s})(1 + r/r_{\rm s})^2}\Theta(r_{\rm vir} - r), \tag{3.92}$$

where Θ is the Heaviside step function.

3.5.2 Mass-concentration relation

The concentration parameter c from the NFW profiles is the link between $\rho_{\rm s}$ and $r_{\rm s}$. That means we can reduce of number of profile parameters to two: mass M and concentration c. Studies show that c is not a constant but varies with M. Therefore, the NFW prescription is
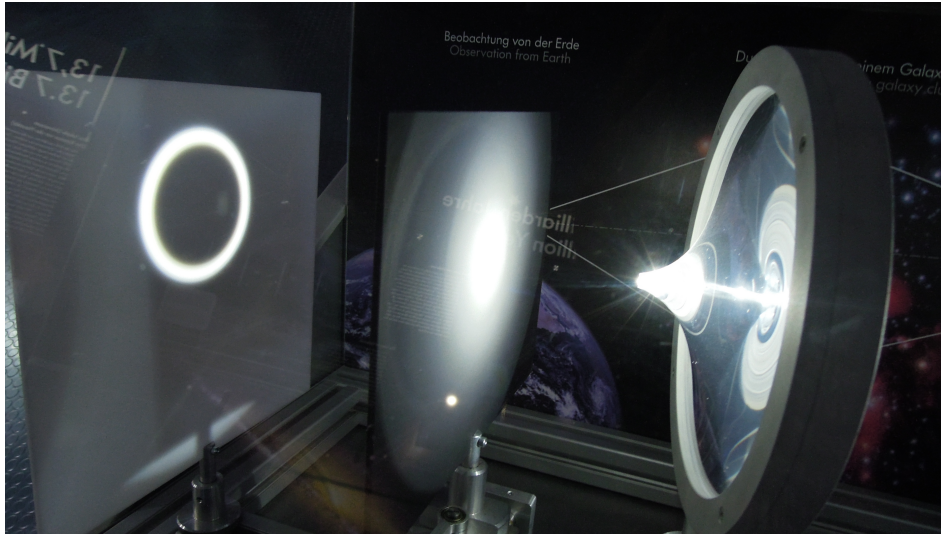
FIGURE 3.5: Creation of an Einstein ring by an optical NFW-like profile! Photo taken at the *Deutsches Museum* in Munich.

always accompanied with a mass-concentration relation (*M-c* relation). Inspired by Bullock et al. (2001), Takada & Jain (2002) proposed a *M-c* relation under a simple form:

$$c(z, M) = \frac{c_0}{1+z} \left(\frac{M}{M_\star}\right)^{-\beta}, \tag{3.93}$$

where $M_\star$ is a pivot mass. In this thesis, I will follow the choice of Takada & Jain (2002) where $M_\star$ satisfies $\delta_c = D(z=0)\sigma(M_\star)$, but it is also common to take $M_\star$ as a constant.

3.5.3 Other profiles

A physically motivated model is the singular isothermal sphere (SIS) profile:

$$\rho_{\mathrm{SIS}}(\boldsymbol{r}) = \frac{\sigma_v^2}{2\pi\mathrm{G}r^2}, \tag{3.94}$$

where σ_v is the 1D velocity dispersion. This profile varies as r^{-2} and fails to describe the inner part of the halos.

Recently, a focus has been put again on the Einasto profile (Einasto 1965). This profile is governed by the exponential function:

$$\rho_{\mathrm{E65}}(\boldsymbol{r}) = \rho_{-2} \exp\left(-\frac{2}{\alpha}\Big((r/r_{-2})^\alpha - 1\Big)\right) \tag{3.95}$$

with a shape parameter α. The subscript $_{-2}$ stands for the position where the density varies as $1/r^2$.

From N-simulations, cosmologists observe that halos are rather triaxial than spherical. On the necessity to model this phenomenon, Jing & Suto (2002) proposed to parametrize r in Eq. (3.86) by

$$r^2 = c^2 \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}\right), \tag{3.96}$$

where x, y, and z are chosen to be aligned with the ellipsoid axes with $a \leq b \leq c$. Jing & Suto (2002) also derived from N-body simulations a numerical fit of the distribution of a/c and b/c.

Finally, Baltz et al. (2009) provided a version of the NFW profiles with a smoothed truncation, which takes form:

$$\rho_{\mathrm{BMO}}(\boldsymbol{r}) = \frac{\rho_{\mathrm{s}}}{(r/r_{\mathrm{s}})(1 + r/r_{\mathrm{s}})^2(1 + r/r_{\mathrm{t}})^n}, \tag{3.97}$$

where r_{t} is the truncation radius and n is the sharpness parameter. A great advantage of this profile is that it models more correctly the regime between the one- and two-halo terms. Baltz et al. (2009) also derived the projected mass from this profile.

Summary

Cosmological structures are the source of weak gravitational lensing. We have reviewed some important mechanisms of structure formation in this chapter.

On large scales, matter structures can be well described by linear perturbation theory, summarized as the evolution of the growth factor.

On small scales, the linear theory is still served as the baseline to understand the structure. Via the spherical collapse model, a connection between the linear and the nonlinear evolution of a virialized object is established, and this connection is used for more detailed studies such as for the mass function.

Two topics have been particularly addressed: the mass function and the halo density profile. Various models have been introduced. These two subjects are essential in the framework of this thesis.

After understanding structure formation, I will focus on developing the theory of gravitational lensing in the next chapter.

Chapter IV

Weak gravitational lensing

Overview

Weak gravitational lensing will be discussed in detail in this chapter. The presentation will start with deriving lensing equations from light deflection. Then, after introducing the convergence and the shear, their expected values from clusters will be introduced. After that, readers will see how the lensing signal can be reconstructed and how cosmological parameters can be extracted from different observables, Gaussian or non-Gaussian. Finally, I will review existing work on weak-lensing peak counts in the literature.

4.1 Light deflection

4.1.1 Different regimes of gravitational lensing

Gravitational lensing is one of the phenomena predicted by general relativity proposed by Einstein (1915). According to this theory, massive objects perturb the spacetime geometry, so that light which always travels on geodesics appears to be bent in a 3D space. For observers who inspect distant objects, images appear at different positions from the original ones, since the direction of the deflected light path converging on the observer is no longer identical to the direct line of sight direction to the source. Therefore, images are subject to transformations such as offset, magnification, and distortion.

This light deflection can also be modelled in classical physics. However, given a point mass, the deflection angle predicted by the classical model is smaller than that in general relativity by a factor 2. This factor was confirmed when the first gravitational lens was observed. In 1919, Dyson et al. (1920) measured the positions of stars when they passed behind the Sun. The observation was feasible thanks to a total solar eclipse, and the results were consistent with Einstein's theory.

Depending on the detection, gravitational lensing can be separated into three regimes. First, when light rays may take more than one path from the source to the observer, the deflection usually results in multiple images or extreme shapes such as arcs and rings. Lensed objects can be easily identified. In this case, the phenomenon is recognized as *strong lensing*. An example of strong lenses is shown in Fig. 4.1. Readers can see that the blue background galaxy has been lensed into at least two arc-like images, forming a near-perfect ring.
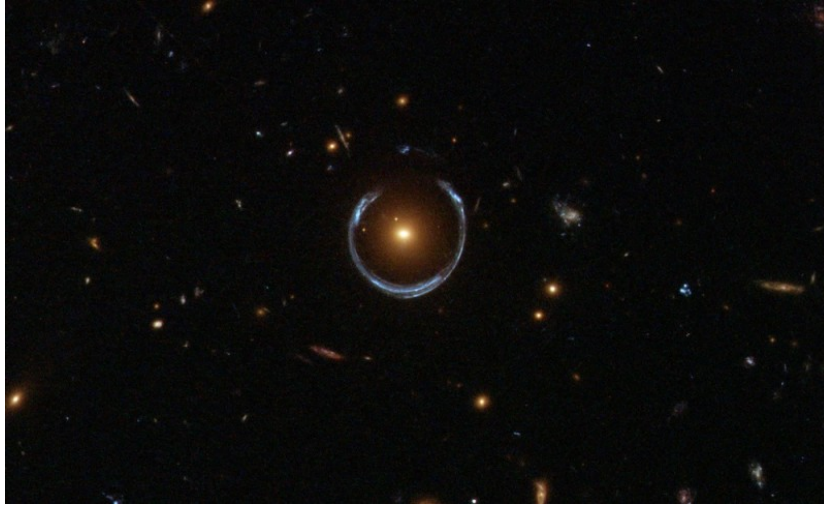
FIGURE 4.1: Image taken by the Hubble Space Telescope. This is a strong lens named LRG 3-757. The blue background galaxy is lensed and is split into multiple images. This ring-like pattern is called Einstein ring. (Source: ESA-Hubble/NASA)
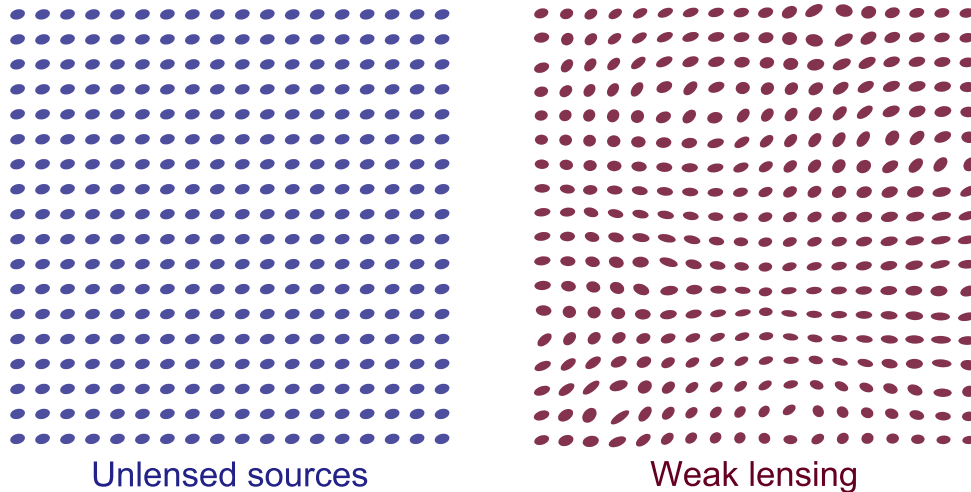
FIGURE 4.2: Illustration of the weak-lensing effect.

The second regime is *weak lensing* (WL). Figure 4.2 shows an exaggerated illustration. In this regime, one can not tell whether images are distorted or not on individual objects. However, by measuring precisely image shapes and by using statistics, signals can still be identified. Cosmologists have started to report significant WL measurements since the 90s: first by cluster lensing (e.g. Bonnet et al. 1994; Luppino & Kaiser 1997), then by galaxy-galaxy lensing (e.g. Brainerd et al. 1996; Fischer et al. 2000), and finally by lensing from large-scale structures (cosmic shear, Bacon et al. 2000; Van Waerbeke et al. 2000; Wittman et al. 2001). Since then, WL has been considered as a tool for probing cosmology, especially for measuring Ω_{m}, σ_8, and w_0^{de}. Recently, the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) has set a milestone by providing interesting WL cosmological constraints. Many important probes are still ongoing, e.g. the Kilo-Degree Survey (KiDS), the Dark Energy Survey (DES), the Subaru Hyper Suprime-Cam (HSC) survey, and the Javalambre Physics of the Accelerating Universe Astrophysical Survey (J-PAS). Furthermore, we look forward in the future to very large lensing data sets from the Euclid mission, the Large Synoptic Survey Telescope (LSST),
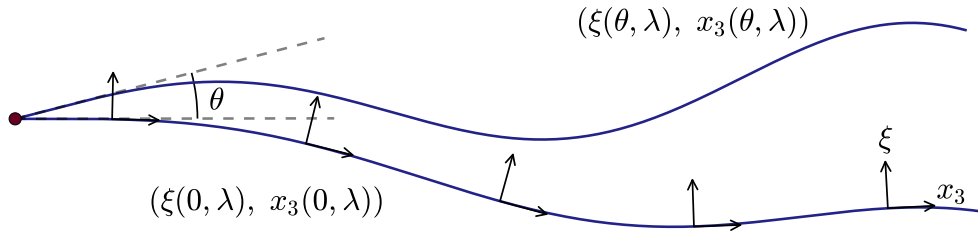
FIGURE 4.3: Illustration of the geodesic deviation. The bottom light ray is chosen as the reference; the top light ray is traced with an exaggerated separation angle θ. Both are parametrized by the affine parameter λ. The space coordinates are decomposed in to a normal part $\boldsymbol{\xi}$ and a tangential part x_3. On the reference line, we have naturally $\boldsymbol{\xi}(0, \lambda) = (0, 0)$.

and the Wide Field Infrared Survey Telescope (WFIRST). To enhance statistical power, these surveys require very precise shape measurements. The objective is to reach the precision of 2×10^{-4} on the average ellipticity (Laureijs et al. 2011), which is an order of magnitude smaller than the deviation of the Earth from a perfect sphere.

Last, if the lens mass is as low as a planet or a star, we enter the *microlensing* regime. In this case, distortion are not measurable, but the difference on apparent magnitude is. Recently, this phenomenon is widely used for searching for exoplanets.

4.1.2 Geodesic deviation equation

In general relativity, the trajectory of light is defined by null geodesics. Therefore, given any metric, light deflection from gravitational lensing can be derived from the geodesic equation. Following Seitz et al. (1994) and Bartelmann (2010), let us select a right ray as the reference geodesic, parametrized by an affine parameter λ. Consider a nearby light ray from a direction characterized by a Cartesian vector $\boldsymbol{\theta} = (\theta_1, \theta_2)$ defining a position on the sky, we define $\boldsymbol{\xi}(\boldsymbol{\theta}, \lambda) = (x_1, x_2)$ as the physical separation vector between two geodesics, i.e. the transverse components which are perpendicular to the tangential direction x_3 of the reference ray in the physical 3D space (Fig. 4.3). Then, the geodesic deviation equation, which characterizes the evolution of $\boldsymbol{\xi}$, leads to

$$\frac{\mathrm{d}^2 \boldsymbol{\xi}(\boldsymbol{\theta}, \lambda)}{\mathrm{d}\lambda^2} = \mathcal{T}(x_1, x_2, x_3)\boldsymbol{\xi}(\boldsymbol{\theta}, \lambda), \tag{4.1}$$

where $\mathcal{T}$ is the *optical tidal matrix*. This matrix can be decomposed into two terms:

$$\mathcal{T} = \mathcal{T}_{\mathrm{bg}} + \mathcal{T}_{\mathrm{cl}}, \tag{4.2}$$

where the background term $\mathcal{T}_{\mathrm{bg}}$ is given by the homogeneous and isotropic Universe, and the clump term $\mathcal{T}_{\mathrm{cl}}$ characterizes local inhomogeneities. Both terms can be expressed by the Riemann, the Ricci, and the Weyl tensors (see e.g. Seitz et al. 1994). Consider now a FLRW metric perturbed by a weak Newtonian gravitational potential Φ. Assuming that Φ is quasi-static and $\Phi \ll \mathrm{c}^2$, the metric is given by

$$\mathrm{d}s^2 = a^2(\tau)\left[-\left(1 + \frac{2\Phi}{\mathrm{c}^2}\right)\mathrm{c}^2\mathrm{d}\tau^2 + \left(1 - \frac{2\Phi}{\mathrm{c}^2}\right)\left(\mathrm{d}w^2 + f_K^2(w)\mathrm{d}\Omega^2\right)\right], \tag{4.3}$$

where a is the scale factor, τ the conformal time, Ω the solid angle, c the light speed, and w and $f_K(w)$ are respectively the comoving radial and transverse distances. In this case, Seitz

et al. (1994) showed that

$$(\mathcal{T}_{\mathrm{bg}})_{ij} = -\frac{4\pi\mathrm{G}}{\mathrm{c}^2}\frac{\bar{\rho}_0}{a^5}\,\delta_{ij}, \tag{4.4}$$

$$(\mathcal{T}_{\mathrm{cl}})_{ij} = -\frac{1}{\mathrm{c}^2 a^2}\left(2\frac{\partial^2}{\partial x_i \partial x_j} + \delta_{ij}\frac{\partial^2}{\partial x_3^2}\right)\phi, \tag{4.5}$$

where $i, j \in \{1, 2\}$, G is the gravitational constant, $\bar{\rho}_0$ the background density at the current time, δ_{ij} the Kronecker delta, and ϕ the reduced Newtonian potential (defined by Eq. 3.38). Here, the reduced potential is needed since $\mathcal{T}_{\mathrm{bg}}$ has already extracted the background part of the potential. Note that x_3 is the line-of-sight direction and x_1 and x_2 are transverse components.

Now, we consider that light is propagated in the *thin-lens approximation*, which means that inhomogeneities are geometrically thin. This allows us to consider 2D clump mass distributions in isolation. In this case, the term $\partial^2\phi/\partial x_3^2$ vanishes due to an integration over x_3 (Seitz et al. 1994). Hereafter, we will always stay in this approximation and consider $\partial^2\phi/\partial x_3^2$ to drop out from Eq. (4.5).

Let us rewrite Eq. (4.1) in the comoving space. Due to the redshift effect, the affine parameter satisfies $\mathrm{d}\lambda = a(\mathrm{cd}t) = a^2(\mathrm{cd}\tau) = a^2\mathrm{d}w$. Let $(X_1, X_2, w) = (x_1/a, x_2/a, x_3/a)$ be the comoving coordinates. Define $\boldsymbol{\Xi} \equiv \boldsymbol{\xi}/a = (X_1, X_2)$ as the separation vector in the comoving space. Denoting $a' = \mathrm{d}a/\mathrm{d}\lambda$, $\boldsymbol{\xi}' = \mathrm{d}\boldsymbol{\xi}/\mathrm{d}\lambda$, we have

$$\frac{\mathrm{d}^2\boldsymbol{\Xi}}{\mathrm{d}w^2} = a^2\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(a\boldsymbol{\xi}' - \boldsymbol{\xi}a'\right) = a^3\boldsymbol{\xi}'' - a^2 a''\boldsymbol{\xi}. \tag{4.6}$$

On the one hand, Eq. (4.1) leads to $\boldsymbol{\xi}'' = (\mathcal{T}_{\mathrm{bg}} + \mathcal{T}_{\mathrm{cl}})\boldsymbol{\xi}$. On the other hand,

$$a'' = \frac{\mathrm{d}^2 a}{\mathrm{d}\lambda^2} = \frac{1}{\mathrm{c}^2 a}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{a}}{a}\right) = \frac{\dot{a}}{\mathrm{c}^2 a}\frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{\dot{a}}{a}\right) = \frac{1}{2\mathrm{c}^2}\frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{\dot{a}}{a}\right)^2, \tag{4.7}$$

where $(\dot{a}/a)^2$ can be given by the Friedmann equation, Eq. (2.6), which results in

$$a'' = -\frac{4\pi\mathrm{G}}{\mathrm{c}^2}\frac{\bar{\rho}_0}{a^4} + \frac{K}{a^3}, \tag{4.8}$$

where K is the spacetime curvature. Using Eqs. (4.4), (4.5), and (4.8), Eq. (4.6) leads to

$$\frac{\mathrm{d}^2 X_i}{\mathrm{d}w^2} = \sum_{j=1,2}\left(-\frac{K}{a}\delta_{ij} - \frac{2a}{\mathrm{c}^2}\frac{\partial^2\phi}{\partial x_i\partial x_j}\right)x_j = -KX_i - \frac{2}{\mathrm{c}^2}\sum_{j=1,2}\left(\frac{\partial^2\phi}{\partial X_i\partial X_j}\right)X_j. \tag{4.9}$$

Since for any 2D differentiable function f, for small X_1 and X_2, we have

$$f(X_1, X_2) - f(0) = \frac{\partial f}{\partial X_1}X_1 + \frac{\partial f}{\partial X_2}X_2, \tag{4.10}$$

applying Eq. (4.10) to $(\partial\phi/\partial X_i)(\,\cdot\,, w)$ yields the comoving geodesic deviation equation under the thin-lens approximation:

$$\frac{\mathrm{d}^2\boldsymbol{\Xi}}{\mathrm{d}w^2} + K\boldsymbol{\Xi} = -\frac{2}{\mathrm{c}^2}\Big[\nabla_\perp\phi(\boldsymbol{\Xi}, w) - \nabla_\perp\Phi(0, w)\Big], \tag{4.11}$$

where $\nabla_\perp = (\partial/\partial X_1, \partial/\partial X_2)$. The dependency of $\boldsymbol{\Xi}$ on w is implicit.

4.1.3 Born approximation and distortion matrix

To solve Eq. (4.11), the technique of the Green's function is used. Given any linear differential operator L and any source function S, the system $L(y(x)) = S(x)$ can be solved by finding a Green's function $G(x, x')$ such that

$$L(G(x, x')) = \delta(x - x'), \tag{4.12}$$

where δ is the Dirac distribution. If G is provided, the particular solution y_{p} can be found by

$$y_{\mathrm{p}}(x) = \int \mathrm{d}x' \; G(x - x')S(x'), \tag{4.13}$$

with homogeneous boundary conditions (i.e. under the form of $y^{(i)}(x_0) = 0$), and the general solution is

$$y(x) = y_{\mathrm{p}}(x) + y_{\mathrm{c}}(x), \tag{4.14}$$

where y_{c} is the solution of the homogeneous equation $Ly = 0$ with the boundary conditions of the original problem.

Applying the above paragraph to Eq. (4.11), readers easily find

$$L(\mathbf{\Xi}) = \frac{\mathrm{d}^2 \mathbf{\Xi}}{\mathrm{d}w^2} + K\mathbf{\Xi}, \tag{4.15}$$

$$S(w) = -\frac{2}{\mathrm{c}^2} \Big[\nabla_\perp \phi(\mathbf{\Xi}, w) - \nabla_\perp \phi(0, w) \Big] \Theta(w), \tag{4.16}$$

where the Heaviside step function Θ is added since $S(w)$ is not defined for $w < 0$, and the boundary conditions are $\mathbf{\Xi}(0) = 0$ and $(\mathrm{d}\mathbf{\Xi}/\mathrm{d}w)|_{w=0} = \boldsymbol{\theta}$. For $w \neq w'$, $\delta(w - w') = 0$, so the Green's function possesses the following form:

$$G(w, w') = \begin{cases} A \cdot f_K(w - w') + B \cdot g_K(w - w') & \text{if } w > w', \\ C \cdot f_K(w - w') + D \cdot g_K(w - w') & \text{if } w < w', \end{cases} \tag{4.17}$$

where A, B, C, and D are constants to be determined and

$$f_K(w) = \begin{cases} \dfrac{1}{\sqrt{K}}\sin(\sqrt{K}w) & \text{if } K > 0, \\ w & \text{if } K = 0, \\ \dfrac{1}{\sqrt{K}}\sinh(\sqrt{K}w) & \text{if } K < 0, \end{cases} \qquad g_K(w) = \begin{cases} \dfrac{1}{\sqrt{K}}\cos(\sqrt{K}w) & \text{if } K > 0, \\ 1 & \text{if } K = 0, \\ \dfrac{1}{\sqrt{K}}\cosh(\sqrt{K}w) & \text{if } K < 0. \end{cases} \tag{4.18}$$

We can see that the function f_K here is the same as Eq. (2.29). The Green function satisfies the homogeneous boundary conditions, i.e. $G(0, w') = (\partial G/\partial w)|_{w=0} = 0$. For a fixed point $w' > 0$, this enforces $C = D = 0$. Then, the order of L is two. This guarantees the continuity at $w = w'$ at zero order and $(\partial G/\partial w)|_{w=w'_+} = (\partial G/\partial w)|_{w=w'_-} + 1$, which implies

$$\begin{cases} A \cdot f_K(0) + B \cdot g_K(0) = 0 \\ A \cdot f'_K(0) + B \cdot g'_K(0) = 1 \end{cases}, \tag{4.19}$$

which results in $A = 1$ and $B = 0$. Therefore, the Green function and the particular solution are

$$G(w, w') = \Theta(w - w')f_K(w - w'), \tag{4.20}$$

$$\boldsymbol{\Xi}_{\mathrm{p}}(w) = -\frac{2}{\mathrm{c}^2} \int_0^w \mathrm{d}w' \, f_K(w - w')\Big[\nabla_\perp \phi(\boldsymbol{\Xi}, w') - \nabla_\perp \phi(0, w')\Big]. \tag{4.21}$$

On the other hand, the solution to the homogeneous equation is also in the form of $A \cdot f_K(w) + B \cdot g_K(w)$. With the boundary condition, we obtain at ease $A = \boldsymbol{\theta}$ and $B = 0$. At the end of the day, the solution to Eq. (4.11) is

$$\boldsymbol{\Xi}(\boldsymbol{\theta}, w) = f_K(w)\boldsymbol{\theta} - \frac{2}{\mathrm{c}^2} \int_0^w \mathrm{d}w' \, f_K(w - w')\Big[\nabla_\perp \phi(\boldsymbol{\Xi}, w') - \nabla_\perp \phi(0, w')\Big]. \tag{4.22}$$

Now we come up with an expression for the separation vector in spite of its recursive definition. This can be resolved by inserting iteratively Eq. (4.22) into itself. The term $\nabla_\perp \phi$ can be developed with regard to $\boldsymbol{\Xi}$ and if the series is truncated at the first term, this is called *Born approximation*, originally used in quantum mechanics for scattering theory. Applying the Born approximation to Eq. (4.22) is simply replacing $\boldsymbol{\Xi}$ at the right-hand side with $f_K(w)\boldsymbol{\theta}$. Physically, it assumes that the potential evaluated on the perturbed light path does not differ substantially from that evaluated on the unperturbed line of sight.

Consider now that $\boldsymbol{\Xi}$ at the right-hand side of Eq. (4.22) is replaced with $f_K(w)\boldsymbol{\theta}$ by the Born approximation. If a light ray is received from the direction $\boldsymbol{\theta}$, the set of $\boldsymbol{\Xi}(\boldsymbol{\theta}, w \geq 0)$ constructs its trajectory, and any point on this path can be a possible position of the source. Say that the source is located at w. Then, the apparent angular position is obviously $\boldsymbol{\theta}$. However, if the inhomogeneity is not present, the unlensed angular position is known. This is $\boldsymbol{\beta}(\boldsymbol{\theta}, w) \equiv \boldsymbol{\Xi}(\boldsymbol{\theta}, w)/f_K(w)$ since $\boldsymbol{\Xi}$ is defined in comoving space. As a result, Eq. (4.22) yields a mapping from the lensed space (observed position $\boldsymbol{\theta}$) to the unlensed space (true position $\boldsymbol{\beta}$). We can characterize this transformation using the 2×2 *distortion matrix* $\mathcal{A}$ defined as

$$\mathcal{A}(\boldsymbol{\theta}, w) \equiv \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \frac{1}{f_K(w)} \frac{\partial \boldsymbol{\Xi}}{\partial \boldsymbol{\theta}}. \tag{4.23}$$

By inserting Eq. (4.22) into Eq. (4.23), this yields

$$\mathcal{A}_{ij}(\boldsymbol{\theta}, w) = \delta_{ij} - \frac{2}{\mathrm{c}^2} \int_0^w \mathrm{d}w' \, \frac{f_K(w - w')f_K(w')}{f_K(w)} \partial_i \partial_j \phi(f_K(w')\boldsymbol{\theta}, w'), \tag{4.24}$$

where $\partial_i \equiv \partial/\partial X_i$ and $i, j \in \{1, 2\}$. It is useful to introduce the *lensing potential* ψ, defined as

$$\psi(\boldsymbol{\theta}, w) \equiv \frac{2}{\mathrm{c}^2} \int_0^w \mathrm{d}w' \, \frac{f_K(w - w')}{f_K(w)f_K(w')} \, \phi(f_K(w')\boldsymbol{\theta}, w'), \tag{4.25}$$

which can be considered as a distance-weighted Newtonian potential. This simplifies the components of the distortion matrix into

$$\mathcal{A}_{ij}(\boldsymbol{\theta}, w) = \delta_{ij} - \partial_i \partial_j \psi(\boldsymbol{\theta}, w), \tag{4.26}$$

where $\partial_i = \partial/\partial \theta_i$.
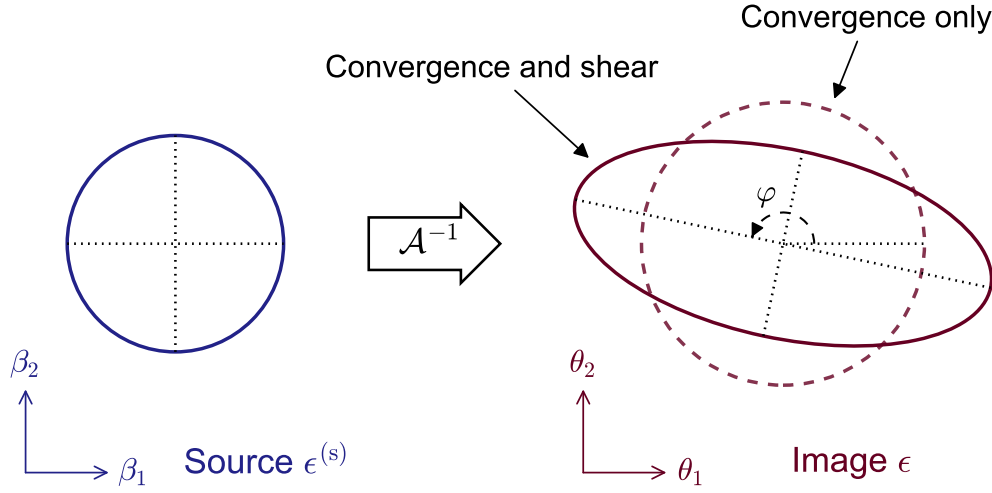
FIGURE 4.4: Illustration of the transformation generated by the distortion matrix.

4.1.4 Convergence and shear

From the previous section, we derive from general relativity the distortion of images from sources due to gravitational lensing. In the linear regime, the distortion is characterized by Eq. (4.26). To interpret the transformation, we parametrize the distortion matrix as

$$\mathcal{A}(\boldsymbol{\theta}, w) = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix}, \tag{4.27}$$

where $g_i(\boldsymbol{\theta}, w) \equiv \gamma_i(\boldsymbol{\theta}, w)/(1 - \kappa(\boldsymbol{\theta}, w))$. Cosmologists usually use the complex notation to denote γ_i and g_i:

$$\gamma \equiv \gamma_1 + \mathrm{i}\gamma_2 = |\gamma| \cdot \mathrm{e}^{2\mathrm{i}\varphi}, \quad g \equiv g_1 + \mathrm{i}g_2 = |g| \cdot \mathrm{e}^{2\mathrm{i}\varphi}, \tag{4.28}$$

where φ is the "rotation angle" that will be discussed later. In this parametrization, the scalar κ is called the *convergence*; the complex number γ is called the *shear* and g the *reduced shear*. It is straightforward to relate κ and γ to ψ by

$$\kappa = \frac{\partial_1^2 + \partial_2^2}{2}\psi, \quad \gamma_1 = \frac{\partial_1^2 - \partial_2^2}{2}\psi, \quad \gamma_2 = \partial_1 \partial_2 \psi. \tag{4.29}$$

Therefore, as we can see, κ and γ are actually related to one another. In the WL regime, we have $|\kappa| < 1$ and $|g| < 1$, and in most cases we often assume $|\kappa| \ll 1$ and $|g| \ll 1$.

Interpretation

What is the interpretation of these quantities? A summary is provided by Fig. 4.4. Consider a circular source S with radius β. The border of this source can be parametrized as $S = (\beta \cos t, \beta \sin t)$, $t \in \mathbb{R}$. Distorted by lensing, the equation of the image I is

$$I = \mathcal{A}^{-1}S = \frac{\beta}{(1 - \kappa)(1 - |g|^2)} \begin{pmatrix} \cos t + |g| \cos(2\varphi - t) \\ \sin t + |g| \sin(2\varphi - t) \end{pmatrix}. \tag{4.30}$$
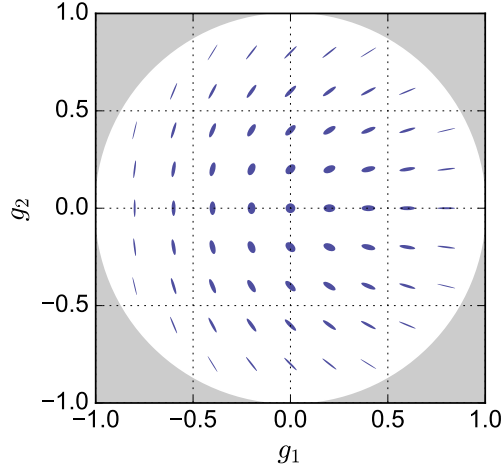
FIGURE 4.5: Illustration of different values of g. Each spot is the image of the same circular source for a given lensing shear. The grey zone corresponds to the strong lensing regime. Readers will also see in Sect. 4.3.1 that this also excludes ellipticities without physical interpretation.

If we rotate the image by $-\varphi$, the rotated image I_{rot} is

$$
I_{\mathrm{rot}} = \begin{pmatrix} \cos(-\varphi) & \sin(-\varphi) \\ -\sin(-\varphi) & \cos(-\varphi) \end{pmatrix} I
$$
$$
= \frac{\beta}{(1-\kappa)(1-|g|^2)} \begin{pmatrix} (1+|g|)\cos(t+\varphi) \\ (1-|g|)\sin(t+\varphi) \end{pmatrix} = \begin{pmatrix} \cos(t+\varphi)\cdot\beta/(1-\kappa-|\gamma|) \\ \sin(t+\varphi)\cdot\beta/(1-\kappa+|\gamma|) \end{pmatrix}. \quad (4.31)
$$

This is the equation of a parametrized ellipse whose semi-axes are respectively $\beta/(1-\kappa\mp|\gamma|)$. If the shear vanishes, the image turns out to be a circle with radius $\beta/(1-\kappa)$. Thus, κ indicates the isotropic part of the transformation. On the other hand, γ contains the distortion information. The norm $|\gamma|$ characterizes the flatness of the ellipse and the rotation angle φ indicates the distortion direction.

Figure 4.5 illustrates the distorted image for different values of the shear, for the same circular source. The ellipses of the same morphological shape are located on the same circle centered at $(g_1, g_2) = (0, 0)$. Note that the rotation angle of the ellipse is not the phase of the shear. Their relation can be obtained easily by reversing Eq. (4.28), which yields

$$
\varphi = \frac{1}{2}\arctan\left(\frac{\gamma_2}{\gamma_1}\right). \quad (4.32)
$$

Another interpretation

Another interpretation exists for the convergence. Let us consider the Gauss theorem applied to gravity:

$$
\int_V \mathrm{d}V \ \nabla\boldsymbol{g} = \oint_S \mathrm{d}S \ \boldsymbol{g}\cdot\boldsymbol{n}, \quad (4.33)
$$

where $\boldsymbol{g} = -(\mathrm{G}\rho/r^2)\cdot\boldsymbol{n}\mathrm{d}V$ is the Newtonian gravitational field and $\boldsymbol{n}$ is the elementary vector normal to an elementary surface $\mathrm{d}S$. The right-hand side of Eq. (4.33) represents the

total gravitational flux passing through a Gauss surface; the left-hand side is the total mass wrapped inside. By taking the Gauss surface as an infinitesimal sphere, Eq. (4.33) becomes $\nabla \boldsymbol{g} = -4\pi \mathrm{G}\rho$. The field is the opposite of the gradient of the potential $\boldsymbol{g} = -\nabla\Phi$, which yields the Poisson equation: $\Delta\Phi = 4\pi \mathrm{G}\rho$.

However, this can not be applied directly in the WL context. If the Gauss surface is submerged in a homogeneous Universe, then the flux vanishes everywhere. In order to account for this effect properly, we need to cancel the density by its background value $\bar{\rho}$. By doing this, we recover Eq. (3.38), which is a Poisson equation for the reduced Newtonian potential ϕ: $\Delta\phi = 4\pi \mathrm{G}(\rho - \bar{\rho})$. Equation (3.39) gives a simplified expression for $\Delta\phi$ in physical coordinates. In comoving space, the 3D Poisson equation is

$$\Delta\phi = \frac{3H_0^2\Omega_{\mathrm{m}}}{2}\frac{\delta}{a}. \tag{4.34}$$

Since $\kappa = 1 - (\mathcal{A}_{11} + \mathcal{A}_{22})/2$, using Eq. (4.24), we find

$$\kappa(\boldsymbol{\theta}, w) = \frac{1}{\mathrm{c}^2}\int_0^w \mathrm{d}w' \, \frac{f_K(w-w')f_K(w')}{f_K(w)}(\Delta - \partial_3^2)\phi(f_K(w')\boldsymbol{\theta}, w'). \tag{4.35}$$

By recalling the thin-lens approximation, $\partial_3^2\Phi$ can be eliminated. Thus, with help of Eq. (4.34), finally we get

$$\kappa(\boldsymbol{\theta}, w) = \frac{3H_0^2\Omega_{\mathrm{m}}}{2\mathrm{c}^2}\int_0^w \mathrm{d}w' \, \frac{f_K(w-w')f_K(w')}{f_K(w)}\frac{\delta\left(f_K(w')\boldsymbol{\theta}, w'\right)}{a(w')}. \tag{4.36}$$

This formula is an integration of the density contrast over the line of sight, weighted by comoving transverse distances and the scale factor. Therefore, *the convergence κ can also be considered as the projected overdensity, or simply the projected mass.* If the source redshift distribution is known as $p(w)$, then it is useful to write Eq. (4.36) into another form which yields the expectation value of κ:

$$\kappa(\boldsymbol{\theta}) = \frac{3H_0^2\Omega_{\mathrm{m}}}{2\mathrm{c}^2}\int_0^{w_{\mathrm{max}}} \mathrm{d}w \, g(w)f_K(w)\frac{\delta\left(f_K(w)\boldsymbol{\theta}, w\right)}{a(w)}, \tag{4.37}$$

where

$$g(w) \equiv \int_w^{w_{\mathrm{max}}} \mathrm{d}w' \, p(w')\frac{f_K(w'-w)}{f_K(w')} \tag{4.38}$$

is the lens efficiency for $p(w)$, and $w_{\mathrm{max}} \equiv w(a=0)$ is the comoving horizon distance.

4.2 Weak lensing by a massive cluster

As we have seen earlier, lensing collects the information of mass distributions along the line of sight. Usually, the signal is insignificant unless a bound massive object, typically a dark matter halo or a cluster of galaxies, has been crossed by the light ray. In this case, we can often neglect the additional lensing by unbound structures, and focus only on the cluster contribution.

In the following, I am going to derive the WL signal in the framework of cluster lensing. This is essential for this thesis work since weak-lensing peak counts depend tightly on cluster population, as we will discuss in Chap. 5.

4.2.1 Mass projection

Consider an isotropic and homogeneous Universe with background density $\bar{\rho}(z) = \rho_{\mathrm{crit}}\Omega_{\mathrm{m}}(1+z)^3$. We put a dark matter halo at w_ℓ (with the corresponding scale factor a_ℓ) and neglect the enhancement of the background density. For a source located at w_{s} (with the corresponding scale factor a_{s}), the convergence κ is provided by Eq. (4.36) as

$$\kappa(\boldsymbol{\theta}, w_{\mathrm{s}}) = \frac{3H_0^2\Omega_{\mathrm{m}}}{2\mathrm{c}^2} \int_0^{w_{\mathrm{s}}} \mathrm{d}w' \; \frac{f_K(w_{\mathrm{s}} - w')f_K(w')}{f_K(w)} \frac{\rho\left(f_K(w')\boldsymbol{\theta}, w'\right) - \bar{\rho}(z(w'))}{\bar{\rho}(z(w'))a(w')}. \tag{4.39}$$

Outside the halo region, the density is equal to the background $\bar{\rho}(z)$, so the integral only extends over the halo region. We neglect here the variation of the scale factor and the comoving distances inside the halo since the halo radius is small compared to w_ℓ in general. In this case, Eq. (4.39) yields

$$\kappa(\boldsymbol{\theta}, w_\ell, w_{\mathrm{s}}) = \kappa_{\mathrm{halo}}(\boldsymbol{\theta}, w_\ell, w_{\mathrm{s}}) - \Delta\kappa(w_\ell, w_{\mathrm{s}}), \tag{4.40}$$

where

$$\kappa_{\mathrm{halo}}(\boldsymbol{\theta}, w_\ell, w_{\mathrm{s}}) = \frac{3H_0^2\Omega_{\mathrm{m}}}{2\mathrm{c}^2} \frac{f_K(w_{\mathrm{s}} - w_\ell)f_K(w_\ell)}{f_K(w_{\mathrm{s}})a_\ell} \int_{\mathrm{halo}} \mathrm{d}w' \; \frac{\rho(f_K(w')\boldsymbol{\theta}, w')}{\bar{\rho}(a_\ell)}, \tag{4.41}$$

$$\Delta\kappa(w_\ell, w_{\mathrm{s}}) = \frac{3H_0^2\Omega_{\mathrm{m}}}{2\mathrm{c}^2} \frac{f_K(w_{\mathrm{s}} - w_\ell)f_K(w_\ell)}{f_K(w_{\mathrm{s}})a_\ell} \int_{\mathrm{halo}} \mathrm{d}w'. \tag{4.42}$$

By analyzing the order of magnitude, we can neglect $\Delta\kappa$: the integral is roughly $\sim (3H_0^2/2\mathrm{c}^2) \cdot R_{\mathrm{vir}} \cdot f(w_\ell)$, where R_{vir} is the comoving quantity which corresponds to the physical virial radius; the factor $3H_0^2/2\mathrm{c}^2$ is $\sim 10^{-8} \; h^2/\mathrm{Mpc}^2$; taking $R_{\mathrm{vir}} \sim 10 \; \mathrm{Mpc}/h$ and $f(w_\ell) \sim 10^3 \; \mathrm{Mpc}/h$, an optimistic estimation yields $\Delta\kappa \sim 10^{-4}$ which justifies that it is negligible.

Let us now concentrate on κ_{halo}, the convergence computed as the projected mass instead of the density contrast. Define (x_1, x_2, x_3) as the physical coordinates associated to $(f_K(w')\boldsymbol{\theta}, w')$ where x_3 is aligned with the line of sight. We aim to rewrite Eq. (4.41) in physical coordinates. First, by the definition of Ω_{m},

$$\frac{3H_0^2\Omega_{\mathrm{m}}}{2\mathrm{c}^2} = \frac{4\pi\mathrm{G}}{\mathrm{c}^2}\bar{\rho}_0, \quad \text{since } \Omega_{\mathrm{m}} = \frac{8\pi\mathrm{G}}{3H_0^2}\bar{\rho}_0. \tag{4.43}$$

Second, by replacing f_K with D_{A}, readers will find

$$\frac{f_K(w_{\mathrm{s}} - w_\ell)f_K(w_\ell)}{f_K(w_{\mathrm{s}})a_\ell} = \frac{(D_{\mathrm{A}}(w_\ell, w_{\mathrm{s}})/a_{\mathrm{s}}) \cdot (D_{\mathrm{A}}(0, w_\ell)/a_\ell)}{(D_{\mathrm{A}}(0, w_{\mathrm{s}})/a_{\mathrm{s}})a_\ell} = \frac{D_\ell D_{\ell\mathrm{s}}}{D_{\mathrm{s}}a_\ell^2}, \tag{4.44}$$

where D_ℓ is the angular diameter distance between the lens and the observer, D_{s} between the source and the observer, and $D_{\ell\mathrm{s}}$ between the lens and the source. Third, $\mathrm{d}w' = \mathrm{d}x_3/a_\ell$. Finally, the matter density varies as the inverse cube of the scale factor, $\bar{\rho}(a_\ell) = \bar{\rho}_0 a_\ell^{-3}$. Putting all together, we get

$$\kappa_{\mathrm{halo}}(\boldsymbol{\theta}, w_\ell, w_{\mathrm{s}}) = \frac{\Sigma_{\mathrm{halo}}(\boldsymbol{\theta})}{\Sigma_{\mathrm{crit}}(w_\ell, w_{\mathrm{s}})}, \tag{4.45}$$

where Σ_{halo} is the projected surface density of the halo:

$$\Sigma_{\mathrm{halo}}(\boldsymbol{\theta}) \equiv \int \mathrm{d}x_3 \; \rho_{\mathrm{halo}}(D_\ell\boldsymbol{\theta}, x_3), \tag{4.46}$$
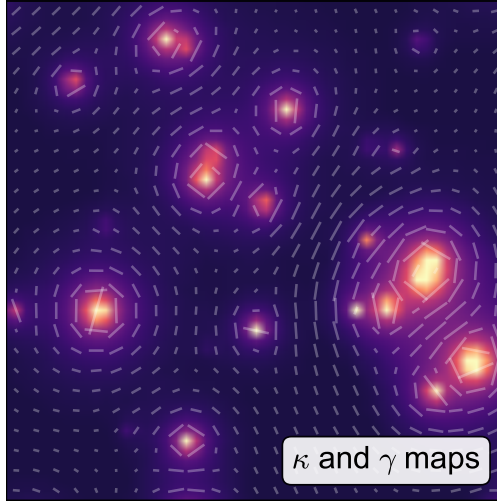
FIGURE 4.6: Illustration of the tangential shear. The colored map represents the convergence, while the "whisker plot" stand for the shear. The length and the orientation of the "whiskers" are decided respectively by the norm and the rotation angle of the tangential shear.

and Σ_{crit} is the *critical surface density*:

$$\Sigma_{\mathrm{crit}}(w_\ell, w_{\mathrm{s}}) \equiv \frac{\mathrm{c}^2}{4\pi \mathrm{G}} \frac{D_{\mathrm{s}}}{D_\ell D_{\ell \mathrm{s}}}. \tag{4.47}$$

In the literature, one will often find that Eq. (4.45) is used rather than Eq. (4.36) for deriving the convergence from a cluster. However, one can not forget the condition in which Eq. (4.45) is valid: *the density outside the halo is not zero but equal to the background value.* Thus, if one considers a lightcone populated with halos, the space which is not occupied by halos is actually **not** empty! If one uses Eq. (4.45) to compute κ in this case, it needs to be corrected, since the real background mass has been enhanced and this enhancement may not be negligible.

4.2.2 Tangential shear profile

In gravitational lensing, sources situated near a point mass will be distorted tangentially to the impact vector. A direct consequence are arc-like features surrounding strong gravitational lenses. In WL, this tangential stretch can still be observed and can be used for cluster identification.

In an ideal scenario where the cluster is perfectly spherical, the rotation angle of the shear will align with the tangential direction and the norm varies depending on the angular distance between the lens and the source. To extract the stretch information, we only have to choose the tangential direction as the reference axis. This depends on the relative position between the cluster and the galaxy. Let ϕ be the phase angle of the position vector of the source pointed from the lens. The *tangential shear* γ_+ and the *cross shear* $\gamma_\times$ components are

$$\gamma_+ \equiv -\mathrm{Re}\left[\gamma \mathrm{e}^{-2\mathrm{i}\phi}\right], \quad \gamma_\times \equiv -\mathrm{Im}\left[\gamma \mathrm{e}^{-2\mathrm{i}\phi}\right]. \tag{4.48}$$

Figure 4.6 shows a convergence map overlapping with a shear map, represented by a "whisker" plot. Readers can easily figure out that the minus sign and $\exp(-2\mathrm{i}\phi)$ account for a rotation by $\pi/2 - \phi$. Similar definitions also exist for the ellipticity ϵ_+ and $\epsilon_\times$ that we will see later.

In more general cases, azimuthally averaging over the tangential shear helps recover the mass enclosed inside the corresponding circular region. This introduces the *tangential shear profile* as $\gamma_+(\theta)$. Given any reference point, the tangential shear profile can be determined from the convergence inside the disk of radius θ without the need of circular symmetry. To establish this relation, let us start with applying the 2D Gauss theorem to $\nabla\psi$ on a disk of radius θ:

$$\theta \int_0^{2\pi} \mathrm{d}\phi \ \nabla\psi \cdot \boldsymbol{n} = \int_{\theta'=0}^{\theta} \int_{\phi=0}^{2\pi} \theta'\mathrm{d}\theta'\mathrm{d}\phi \ \nabla \cdot \nabla\psi. \tag{4.49}$$

From Eq. (4.29), $\nabla \cdot \nabla\psi = \Delta\psi = 2\kappa$. Also, $\nabla\psi \cdot \boldsymbol{n} = \partial\psi/\partial\theta$, so

$$\theta \cdot \left\langle \frac{\partial\psi}{\partial\theta} \right\rangle_{\mathrm{c}} (\theta) = 2 \int_0^{\theta} \theta'\mathrm{d}\theta' \ \langle\kappa\rangle_{\mathrm{c}}(\theta'), \tag{4.50}$$

where

$$\langle f\rangle_{\mathrm{c}}(\theta) \equiv \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\phi \ f(\theta, \phi), \tag{4.51}$$

is the circular average of any quantity f defined in a 2D space. By deriving Eq. (4.50) with regard to θ, we obtain

$$\left\langle \frac{\partial\psi}{\partial\theta} \right\rangle_{\mathrm{c}} (\theta) + \theta \cdot \left\langle \frac{\partial^2\psi}{\partial\theta^2} \right\rangle_{\mathrm{c}} (\theta) = 2\theta \cdot \langle\kappa\rangle_{\mathrm{c}}(\theta). \tag{4.52}$$

The first term at the left-hand side of Eq. (4.52) can be replaced with Eq. (4.50). For the second term, $\partial^2\psi/\partial\theta^2$ is actually identical to $\kappa - \gamma_+$. This can be justified easily when $\phi = 0$ since in this case, $\gamma_+ = -\gamma_1$, and the general cases can be derived by rotation. The resulting formula is

$$\frac{2}{\theta} \int_0^{\theta} \theta'\mathrm{d}\theta' \ \langle\kappa\rangle_{\mathrm{c}}(\theta') - \theta \cdot \langle\gamma_+\rangle_{\mathrm{c}}(\theta) = \theta \cdot \langle\kappa\rangle_{\mathrm{c}}(\theta). \tag{4.53}$$

Defining $\overline{f}(< \theta)$ as the mean of any quantity f over a disk of radius θ:

$$\pi\theta^2 \cdot \overline{f}(< \theta) \equiv \int_{\mathrm{disk}} \mathrm{d}\boldsymbol{\theta}' \ f(\boldsymbol{\theta}') = 2\pi \int_0^{\theta} \theta'\mathrm{d}\theta' \ \langle f\rangle_{\mathrm{c}}(\theta'), \tag{4.54}$$

we obtain finally

$$\langle\gamma_+\rangle_{\mathrm{c}}(\theta) = \overline{\kappa}(< \theta) - \langle\kappa\rangle_{\mathrm{c}}(\theta). \tag{4.55}$$

If the reference point is the center of a dark matter halo with the projected mass Σ_{halo}, from Eq. (4.45), the derived tangential shear profile is

$$\langle\gamma_{+,\mathrm{halo}}\rangle_{\mathrm{c}}(\theta) = \frac{\overline{\Sigma}_{\mathrm{halo}}(< \theta) - \langle\Sigma_{\mathrm{halo}}\rangle_{\mathrm{c}}(\theta)}{\Sigma_{\mathrm{crit}}}. \tag{4.56}$$

4.2.3   Example: NFW profiles

In this section, we show the convergence and the tangential shear profile for the NFW profiles (Navarro et al. 1996, 1997), for both truncated and non-truncated cases. If the profile is spherically symmetric, Eq. (4.46) can be rewritten into

$$\Sigma_{\mathrm{halo}}\left(\theta = \frac{x_1}{D_\ell}\right) = \int \mathrm{d}x_3 \ \rho_{\mathrm{halo}}\left(\sqrt{x_1^2 + x_3^2}\right). \tag{4.57}$$
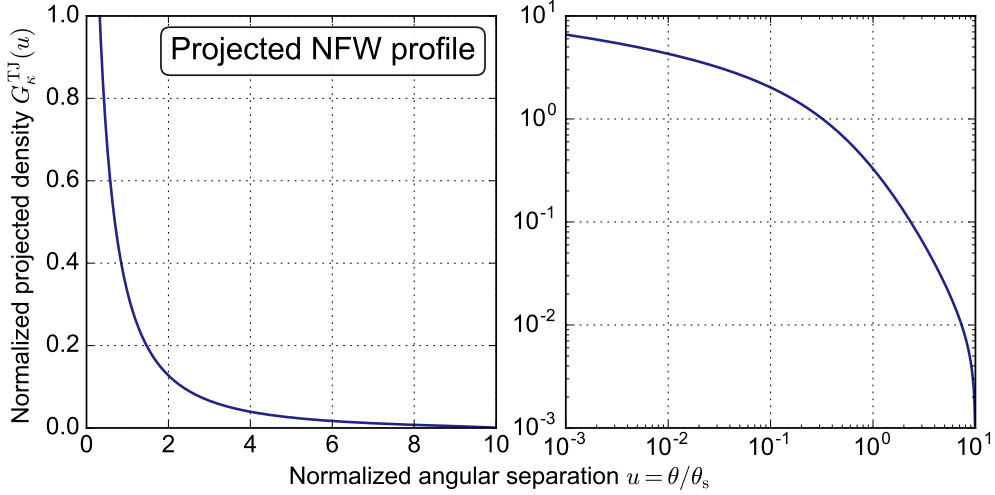
FIGURE 4.7: G_κ^{TJ} as a function of u. This is the projection of the truncated NFW profile shown in normalized coordinates. The concentration is set to 10 as an example. The left panel shows the profile in linear space, and the right panel represents the same plot in log space.

Let $u = x_1/r_\mathrm{s}$ and $v = x_3/r_\mathrm{s}$ be the coordinates normalized by the scale radius r_s of the NFW profiles. The projected mass becomes

$$\Sigma_\mathrm{halo}\left(\theta = \frac{r_\mathrm{s}}{D_\ell}u\right) = \int_{-r_\mathrm{s}v_\mathrm{max}}^{r_\mathrm{s}v_\mathrm{max}} \mathrm{d}(r_\mathrm{s}v) \; \frac{\rho_\mathrm{s}}{\sqrt{u^2+v^2}\cdot(1+\sqrt{u^2+v^2})^2}$$
$$= 2\rho_\mathrm{s} r_\mathrm{s} \int_0^{v_\mathrm{max}} \frac{\mathrm{d}v}{\sqrt{u^2+v^2}\cdot(1+\sqrt{u^2+v^2})^2}, \tag{4.58}$$

with

$$v_\mathrm{max} = \begin{cases} +\infty & \text{for the exact NFW profiles,} \\ \sqrt{c^2-u^2} & \text{for the profiles truncated at } r_\mathrm{vir}, \end{cases} \tag{4.59}$$

where ρ_s is the NFW characteristic density, r_vir the virial radius, and c the concentration parameter. Thus, the problem is reduced to the calculation of the dimensionless projected mass $G_\kappa(u)$:

$$G_\kappa(u) \equiv \int_0^{v_\mathrm{max}} \frac{\mathrm{d}v}{\sqrt{u^2+v^2}\cdot(1+\sqrt{u^2+v^2})^2}. \tag{4.60}$$

Following Wright & Brainerd (2000) and Takada & Jain (2003b), the projected masses for the original NFW profiles (Eq. 3.86) and the truncated ones (labelled `TJ`, Eq. 3.92) are

$$G_\kappa^{\mathrm{NFW}}(u) = \begin{cases} \text{not defined} & \text{if } u = 0, \\[2mm] -\dfrac{1}{1-u^2} + \dfrac{1}{(1-u^2)^{3/2}}\,\mathrm{arcosh}\left(\dfrac{1}{u}\right) & \text{if } 0 < u < 1, \\[2mm] \dfrac{1}{3} & \text{if } u = 1, \\[2mm] \dfrac{1}{u^2-1} - \dfrac{1}{(u^2-1)^{3/2}}\,\arccos\left(\dfrac{1}{u}\right) & \text{if } u > 1, \end{cases} \tag{4.61}$$

and

$$
G_\kappa^{\mathrm{TJ}}(u) = \begin{cases}
\text{not defined} & \text{if } u = 0, \\[2mm]
-\dfrac{1}{1-u^2}\dfrac{\sqrt{c^2-u^2}}{c+1} + \dfrac{1}{(1-u^2)^{3/2}}\,\mathrm{arcosh}\left[\dfrac{u^2+c}{u(c+1)}\right] & \text{if } 0 < u < 1, \\[4mm]
\dfrac{\sqrt{c^2-1}}{c+1}\cdot\dfrac{c+2}{3(c+1)} & \text{if } u = 1, \\[4mm]
\dfrac{1}{u^2-1}\dfrac{\sqrt{c^2-u^2}}{c+1} - \dfrac{1}{(u^2-1)^{3/2}}\,\mathrm{arccos}\left[\dfrac{u^2+c}{u(c+1)}\right] & \text{if } 1 < u \le c, \\[4mm]
0 & \text{if } u > c.
\end{cases}
\tag{4.62}
$$

These functions are not defined at $u = 0$ since the NFW profiles diverge to infinity when $r \to 0$. Note that $\mathrm{arcosh}(1/u) = 2\,\mathrm{artanh}\sqrt{\frac{1-u}{1+u}}$ for $0 < u < 1$ and $\mathrm{arccos}(1/u) = 2\arctan\sqrt{\frac{u-1}{u+1}}$ for $u > 1$. Therefore, gathering Eqs. (4.45), (4.47), (4.58), and (4.60), the convergence projected from the NFW profiles is

$$
\kappa_{\mathrm{halo}}(\boldsymbol{\theta}) = \frac{2\rho_s r_s}{\Sigma_{\mathrm{crit}}}\cdot G_\kappa\left(\frac{\theta}{\theta_s}\right) = \frac{4\pi\mathrm{G}}{\mathrm{c}^2}\frac{D_\ell D_{\ell s}}{D_s}\cdot\frac{Mfc^2}{2\pi r_{\mathrm{vir}}^2}\cdot G_\kappa\left(\frac{c\theta}{\theta_{\mathrm{vir}}}\right),
\tag{4.63}
$$

where $\theta_s \equiv r_s/D_\ell$, $\theta_{\mathrm{vir}} \equiv r_{\mathrm{vir}}/D_\ell$, M is the halo mass, and f is given by Eq. (3.89). The equality $2\rho_s r_s = (Mfc^2)/(2\pi r_{\mathrm{vir}}^2)$ is useful for implementation.

From Eq. (4.56), we can derive the tangential shear profile. The profiles are spherical, so we simplify the notation $\langle\Sigma_{\mathrm{halo}}\rangle_c(\theta)$ to $\Sigma_{\mathrm{halo}}(\theta)$ and $\langle\gamma_{+,\mathrm{halo}}\rangle_c(\theta)$ to $\gamma_{+,\mathrm{halo}}(\theta)$. Inserting Eq. (4.58) into Eq. (4.56), we obtain

$$
\gamma_{+,\mathrm{halo}}(\theta) = \frac{2\rho_s r_s}{\Sigma_{\mathrm{crit}}}\cdot G_\gamma\left(\frac{\theta}{\theta_s}\right),
\tag{4.64}
$$

with

$$
G_\gamma(u) = \overline{G}_\kappa(< u) - G_\kappa(u) = \frac{2}{u^2}\int_0^u u'\mathrm{d}u'\, G_\kappa(u') - G_\kappa(u).
\tag{4.65}
$$

The G_γ for the original NFW profiles and the truncated ones (labelled TJ) are given by Wright & Brainerd (2000) and Takada & Jain (2003c) as

$$
G_\gamma^{\mathrm{NFW}}(u) = \begin{cases}
\text{not defined} & \text{if } u = 0, \\[2mm]
\dfrac{1}{1-u^2} + \dfrac{2}{u^2}\ln\dfrac{u}{2} + \dfrac{1}{\sqrt{1-u^2}}\left(\dfrac{-1}{1-u^2} + \dfrac{2}{u^2}\right)\mathrm{arcosh}\left(\dfrac{1}{u}\right) & \text{if } 0 < u < 1, \\[4mm]
\dfrac{5}{3} + 2\ln\dfrac{1}{2} & \text{if } u = 1, \\[4mm]
\dfrac{-1}{u^2-1} + \dfrac{2}{u^2}\ln\dfrac{u}{2} + \dfrac{1}{\sqrt{u^2-1}}\left(\dfrac{1}{u^2-1} + \dfrac{2}{u^2}\right)\mathrm{arccos}\left(\dfrac{1}{u}\right) & \text{if } u > 1,
\end{cases}
\tag{4.66}
$$

and

$$
G_\gamma^{\mathrm{TJ}}(u) = \begin{cases}
\text{not defined} & \text{if } u = 0, \\[2ex]
\begin{aligned}
&\frac{1}{u^2(c+1)}\left[\frac{2-u^2}{1-u^2}\sqrt{c^2-u^2}-2c\right]+\frac{2}{u^2}\ln\left[\frac{u(c+1)}{c+\sqrt{c^2-u^2}}\right] \\
&\qquad +\frac{2-3u^2}{u^2(1-u^2)^{3/2}}\operatorname{arcosh}\left[\frac{u^2+c}{u(c+1)}\right]
\end{aligned} & \text{if } 0 < u < 1, \\[4ex]
\frac{1}{3(c+1)}\left[\frac{11c+10}{c+1}\sqrt{c^2-1}-6c\right]+2\ln\left[\frac{c+1}{c+\sqrt{c^2-1}}\right] & \text{if } u = 1, \\[3ex]
\begin{aligned}
&\frac{1}{u^2(c+1)}\left[\frac{2-u^2}{1-u^2}\sqrt{c^2-u^2}-2c\right]+\frac{2}{u^2}\ln\left[\frac{u(c+1)}{c+\sqrt{c^2-u^2}}\right] \\
&\qquad -\frac{2-3u^2}{u^2(u^2-1)^{3/2}}\arccos\left[\frac{u^2+c}{u(c+1)}\right]
\end{aligned} & \text{if } 1 < u \le c, \\[4ex]
\frac{2}{fu^2} & \text{if } u > c.
\end{cases}
\tag{4.67}
$$

4.3 Extraction of cosmological information

4.3.1 From shape to shear

Earlier, we have used the convergence κ and the shear γ to characterize the WL effect. To estimate these lensing quantities, we need to measure the difference between galaxies' lensed and unlensed shapes. Unfortunately, we can not measure the unlensed form of galaxies. However, if cosmologists assume the isotropy of the galaxy intrinsic orientation, then the average shape over some galaxies should be circular. If now we detect a significant elliptical average, then we know that images have been lensed. From this "elliptical average" we should also be able to measure the lensing signal. To do this, we first need to define what the ellipticity is.

Ellipticity definition

The definition with moments is adopted for ellipticity in this work. Let $\boldsymbol{\theta}$ be 2D Cartesian coordinates for angular position, and $I(\boldsymbol{\theta})$ be the surface brightness (brightness density) of an image at $\boldsymbol{\theta}$, and we assume that the image is not contaminated by other images. We define the center of the image as

$$
\bar{\boldsymbol{\theta}} \equiv \frac{\int \mathrm{d}^2\boldsymbol{\theta}\,\omega(I(\boldsymbol{\theta}))I(\boldsymbol{\theta})\cdot\boldsymbol{\theta}}{\int \mathrm{d}^2\boldsymbol{\theta}\,\omega(I(\boldsymbol{\theta}))I(\boldsymbol{\theta})},
\tag{4.68}
$$

where $\omega(I)$ is a weight function of suitable choice, which can be used to define the image border. Then, the second-order moments of the image are

$$
Q_{ij} \equiv \frac{\int \mathrm{d}^2\boldsymbol{\theta}\,\omega(I(\boldsymbol{\theta}))I(\boldsymbol{\theta})\cdot(\theta_i-\bar{\theta}_i)(\theta_j-\bar{\theta}_j)}{\int \mathrm{d}^2\boldsymbol{\theta}\,\omega(I(\boldsymbol{\theta}))I(\boldsymbol{\theta})}.
\tag{4.69}
$$

With this definition, the trace of Q describes the image size, whereas the traceless part contains the distortion. From Q_{ij}, we define the *ellipticity* ϵ as

$$\epsilon \equiv \epsilon_1 + i\epsilon_2, \quad \text{and} \quad \epsilon_1 \equiv \frac{Q_{11} - Q_{22}}{Q_{11} + Q_{22} + 2\sqrt{Q_{11}Q_{22} - Q_{12}^2}},$$

$$\epsilon_2 \equiv \frac{2Q_{12}}{Q_{11} + Q_{22} + 2\sqrt{Q_{11}Q_{22} - Q_{12}^2}}. \tag{4.70}$$

Note that similar to the shear, the ellipticity is described by a complex number whose norm is always smaller than one, since

$$|\epsilon|^2 = \frac{Q_{11} + Q_{22} - 2\sqrt{\Delta}}{Q_{11} + Q_{22} + 2\sqrt{\Delta}}, \quad \Delta \equiv Q_{11}Q_{22} - Q_{12}^2. \tag{4.71}$$

In analogy, we can define the intrinsic ellipticity in the unlensed space (or the source space). Following the same notation as Sect. 4.1.3, we denote $\boldsymbol{\beta}$ the coordinates of angular position in the source space. The corresponding second-order moments lead to

$$Q_{ij}^{(s)} \equiv \frac{\int d^2\boldsymbol{\beta}\ \omega(I^{(s)}(\boldsymbol{\beta}))I^{(s)}(\boldsymbol{\beta}) \cdot (\beta_i - \bar{\beta}_i)(\beta_j - \bar{\beta}_j)}{\int d^2\boldsymbol{\beta}\ \omega(I^{(s)}(\boldsymbol{\beta}))I^{(s)}(\boldsymbol{\beta})}, \tag{4.72}$$

where $I^{(s)}(\boldsymbol{\beta})$ is the surface brightness distribution in the source space, which satisfies the equality $I^{(s)}(\boldsymbol{\beta}) = I(\boldsymbol{\theta})$ (see Sect. 4.3.2 for details). The intrinsic ellipticity $\epsilon^{(s)}$ is defined as

$$\epsilon^{(s)} \equiv \frac{Q_{11}^{(s)} - Q_{22}^{(s)} + 2iQ_{12}^{(s)}}{Q_{11}^{(s)} + Q_{22}^{(s)} + 2\sqrt{Q_{11}^{(s)}Q_{22}^{(s)} - Q_{12}^{(s)2}}}, \tag{4.73}$$

By the definition of $\mathcal{A}$, $\beta_i - \bar{\beta}_i = \sum_{j=1,2} \mathcal{A}_{ij} \cdot (\theta_j - \bar{\theta}_j)$, so

$$(\beta_i - \bar{\beta}_i)(\beta_j - \bar{\beta}_j) = \sum_{k=1,2}\sum_{\ell=1,2} \mathcal{A}_{ik}(\theta_k - \bar{\theta}_k)(\theta_\ell - \bar{\theta}_\ell)\mathcal{A}_{\ell j}^T. \tag{4.74}$$

This results in $Q^{(s)} = \mathcal{A}Q\mathcal{A}^T = \mathcal{A}Q\mathcal{A}$. After some simple algebra (Seitz & Schneider 1997), the intrinsic ellipticity is related to the observed one by the reduced shear g as

$$\epsilon^{(s)} = \begin{cases} \dfrac{\epsilon - g}{1 - g^*\epsilon} & \text{if } |g| \leq 1, \\[3mm] \dfrac{1 - g\epsilon^*}{\epsilon^* - g^*} & \text{if } |g| \geq 1. \end{cases} \tag{4.75}$$

Shear estimator

Now, the question becomes: how to estimate g? More precisely, if we average ϵ locally, what would be the expected value if images are lensed and if $\epsilon^{(s)}$ is statistically isotropic? To figure this out, let us inverse Eq. (4.75) into

$$\epsilon = \begin{cases} \dfrac{\epsilon^{(s)} + g}{1 + g^*\epsilon^{(s)}} & \text{if } |g| \leq 1, \\[3mm] \dfrac{1 + g\epsilon^{(s)*}}{\epsilon^{(s)*} + g^*} & \text{if } |g| \geq 1. \end{cases} \tag{4.76}$$
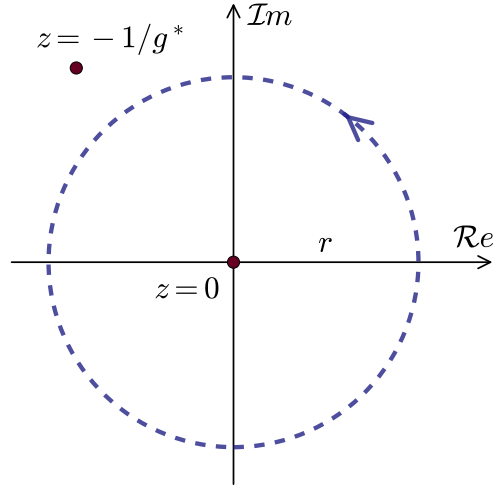
FIGURE 4.8: Illustration of the residue theorem. For the case of $|g| \leq 1$, only one pole out of two stays inside the integration contour.

Let $\epsilon^{(s)} = re^{2i\varphi}$. For all function $f(\epsilon^{(s)})$, the weighted average of f with regard to the intrinsic ellipticity distribution is defined as

$$\langle f \rangle \equiv \frac{\int r \mathrm{d}r \mathrm{d}\varphi \; P(r) \cdot f}{\int r \mathrm{d}r \mathrm{d}\varphi \; P(r)}, \tag{4.77}$$

where $P(r)$ is the probability distribution of $\epsilon^{(s)}$. The dependency of P on φ has been omitted because of isotropy. If ϵ is lensing-free, then the isotropic statement will yield $\langle \epsilon \rangle = \langle \epsilon^{(s)} \rangle = 0$. In general cases, if $|g| \leq 1$, the numerator of Eq. (4.77) is

$$\int r \mathrm{d}r \mathrm{d}\varphi \; P(r) \cdot \epsilon = \int_0^1 r \mathrm{d}r \; P(r) \cdot I(r), \tag{4.78}$$

with

$$I(r) = \int_0^\pi \mathrm{d}\varphi \; \frac{re^{2i\varphi} + g}{1 + g^* re^{2i\varphi}}. \tag{4.79}$$

This expression can be integrated in the complex plane using the residue theorem. Fix r and let $z = re^{2i\varphi}$, we obtain $\mathrm{d}z = 2ire^{2i\varphi}\mathrm{d}\varphi = 2iz\mathrm{d}\varphi$, so that

$$I(r) = \frac{1}{2i} \oint \mathrm{d}z \; \frac{z+g}{z(1+g^*z)} = \frac{1}{2i} \oint \mathrm{d}z \left(\frac{g}{z} + \frac{1-|g|^2}{1+g^*z} \right). \tag{4.80}$$

Since $0 \leq r \leq 1$ and $|g| \leq 1$, we have $r \leq 1 \leq |1/g^*|$. There is only one pole $z = 0$ inside the contour (Fig. 4.8). One obtains

$$I(r) = \frac{1}{2i} \oint \mathrm{d}z \; \frac{g}{z} = \pi \cdot g. \tag{4.81}$$

The denominator of Eq. (4.77) leads to $\int r \mathrm{d}r \; \pi P(r)$. Thus, using Eqs. (4.78) and (4.81), we find

$$\langle \epsilon \rangle = g. \tag{4.82}$$

A similar calculation can be done for $|g| \geq 1$. Finally, the expected average of the observed ellipticities is

$$\langle \epsilon \rangle = \begin{cases} g & \text{if } |g| \leq 1, \\ 1/g^* & \text{if } |g| \geq 1. \end{cases} \tag{4.83}$$

Equation (4.83) means that, whatever the intrinsic ellipticity distribution is, as soon as it is isotropic, then *the mean of ellipticity samples yields locally an **unbiased** estimator of the reduced shear*. Cosmologists usually consider the regime where $|\kappa| \ll 1$. In this case, $g \approx \gamma$. The implication is that the convergence is not directly measurable, but the shear is, and we need to use an inversion to retrieve κ.

Let us revise the notion of shape at the end of this section. Shape is a concept which is very ambiguous for a source galaxy. Even if the galaxy image is isolated from other objects, it is not straightforward to define the boundary since the surface brightness decreases continuously from the inner part. Also, even if the boundary is defined, the word "shape" in its common sense only correspond to the case of isophotes. Therefore, it would be better to replace the notion of shape with ellipticity. On the other hand, setting $\epsilon^{(\mathrm{s})} = 0$, Eq. (4.76) yields $\epsilon = g$. This is exactly the transformation described by Eq. (4.30) and Fig. 4.5. Hence, the shear and the ellipticity are mathematically the same[#1], and Fig. 4.5 can also be interpreted as the elliptical fit to images.

4.3.2 Magnification effect

Apart from distortion, WL also magnifies images. By measure theory, the magnification is given by the Jacobian matrix of the change of variables between the source space and the observed space, which is just the distortion matrix. More precisely, the magnification μ is defined as

$$\mu \equiv \left| \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\beta}} \right| = \det \mathcal{A}^{-1} = \frac{1}{(1-\kappa)^2 - |\gamma|^2}. \tag{4.84}$$

In the WL regime, if $|\kappa| \ll 1$ and $|\gamma| \ll 1$, the dominant term in the denominator is $1 - 2\kappa$. Thus, we relate the magnification to the convergence by $\mu \approx 1 + 2\kappa$.

The magnification effect makes angular separations larger if $\kappa > 0$. Consequently, images become larger and more distant from each other. This decreases locally the number density of sources. Meanwhile, the brightness of sources is enhanced, so more galaxies will be detected. In the following, I will construct the mathematical formalism for this twofold effect.

According to the Liouville theorem, the surface brightness density is invariant to the lensing effect, which means that

$$I(\boldsymbol{\theta}) = I^{(\mathrm{s})}(\boldsymbol{\beta}). \tag{4.85}$$

This can be understood in the following way. If an image is magnified by lensing, its apparent angle becomes larger. Meanwhile, lensing converges light so some photons which would have escaped from the observer are deflected and contribute to the total flux. Therefore, the image size and the total flux increase at the same time, and the Liouville theorem tells us that both increase at the same rate.

[#1]Let $G = \{x | x \in \mathbb{C}, |x| \leq 1\}$ be the set of ellipticities and shears whose norm is smaller than 1. Define $*$ as the "lensing operator": $x * y \equiv (x + y)/(1 + x^*y)$. Then $(G, *)$ is a *magma*: $\forall x, y \in G, x * y \in G$.

PhD thesis of Chieh-An Lin

Let $\mathrm{d}\Omega$ and $\mathrm{d}\Omega^{(\mathrm{s})}$ be a small surface fraction of the sky (solid angle) of an image in each space and $\mathrm{d}B$ and $\mathrm{d}B^{(\mathrm{s})}$ the corresponding brightness. The implication of the Liouville theorem is $\mathrm{d}B/\mathrm{d}\Omega = \mathrm{d}B^{(\mathrm{s})}/\mathrm{d}\Omega^{(\mathrm{s})}$. On the other hand, the ratio between two image sizes is the magnification, so

$$\mu = \frac{\mathrm{d}\Omega}{\mathrm{d}\Omega^{(\mathrm{s})}} = \frac{\mathrm{d}B}{\mathrm{d}B^{(\mathrm{s})}}. \tag{4.86}$$

Let $n(B, z)$ be the surface number density of observed galaxies with brightness B at redshift z and $n^{(\mathrm{s})}(B^{(\mathrm{s})}, z)$ the corresponding function in the $\boldsymbol{\beta}$ space. Let $n(B, z)\mathrm{d}B\mathrm{d}\Omega\mathrm{d}z$ be the number counts of galaxies within solid angle $\mathrm{d}\Omega$ and redshift slice $\mathrm{d}z$ with brightness in $[B, B + \mathrm{d}B[$. Let $\mathrm{d}B^{(\mathrm{s})}$ be the corresponding brightness gap in the $\boldsymbol{\beta}$ space and $\mathrm{d}\Omega^{(\mathrm{s})}$ the corresponding spanning area. Then, the number-count conservation leads to the equality

$$n(B, z)\mathrm{d}B\mathrm{d}\Omega\mathrm{d}z = n^{(\mathrm{s})}(B^{(\mathrm{s})}, z)\mathrm{d}B^{(\mathrm{s})}\mathrm{d}\Omega^{(\mathrm{s})}\mathrm{d}z. \tag{4.87}$$

From Eq. (4.86), the brightness B in the lensed space will become B/μ in the $\boldsymbol{\beta}$ space. By integrating between B (resp. $B^{(\mathrm{s})}$) and infinity,

$$\int_B^\infty \mathrm{d}B\ n(B, z) = \frac{1}{\mu} \int_{B^{(\mathrm{s})}=B/\mu}^\infty \mathrm{d}B^{(\mathrm{s})}\ n^{(\mathrm{s})}(B^{(\mathrm{s})}, z). \tag{4.88}$$

Denote $n(> B, z)$ (resp. $n^{(\mathrm{s})}(> B^{(\mathrm{s})}, z)$) as the number density of galaxies with brightness larger than B (resp. $B^{(\mathrm{s})}$) at redshift z. We find the relation of magnification effect (see also Broadhurst et al. 1995):

$$n(> B, z) = \frac{1}{\mu}\ n^{(\mathrm{s})}\left(> \frac{B}{\mu}, z\right). \tag{4.89}$$

Historically, this effect is called *magnification bias* since it biases the probability of quasar detection (Turner 1980; Canizares 1981). As said earlier, the source number counts with regard to brightness are subject to two effects. The preceding factor $1/\mu$ describes the field-of-view distortion: the stretching area increases angular separations to each other; and the inner factor represents the flux amplification. As illustrated by Fig. 4.9, if $\mu > 1$, the first factor diminishes the distribution function globally (number counts decrease); and the second effect shifts the curve toward the right (number counts increase). At the end of the day, the impact of the magnification on galaxy number density is ambiguous!

From Eq. (4.89), we can construct an estimator of μ as follows. By assuming $n^{(\mathrm{s})}(> B) \propto B^{-\alpha}$, if we can properly estimate the unlensed surface density $n^{(\mathrm{s})}(> B)$, e.g. counting galaxies from cluster-free areas, then the magnification μ will become an observable since

$$\frac{n(> B)}{n^{(\mathrm{s})}(> B)} = \mu^{\alpha-1}. \tag{4.90}$$

And the convergence κ is also observable now! Thanks to Eq. (4.84), the estimator of κ is

$$\kappa \approx \frac{1}{2(1-\alpha)}\left(1 - \frac{n(> B)}{n^{(\mathrm{s})}(> B)}\right). \tag{4.91}$$

However, it is very difficult to measure the number density locally, either for the lensed or for unlensed density. A plausible way to overcome this difficulty is stacking. By stacking clusters of similar mass and redshift, the statistic can be enhanced. The stacked number density estimator provides a probe of the convergence profile, which can be combined with the tangential shear profile to constrain halo profiles more tightly (Umetsu et al. 2014).
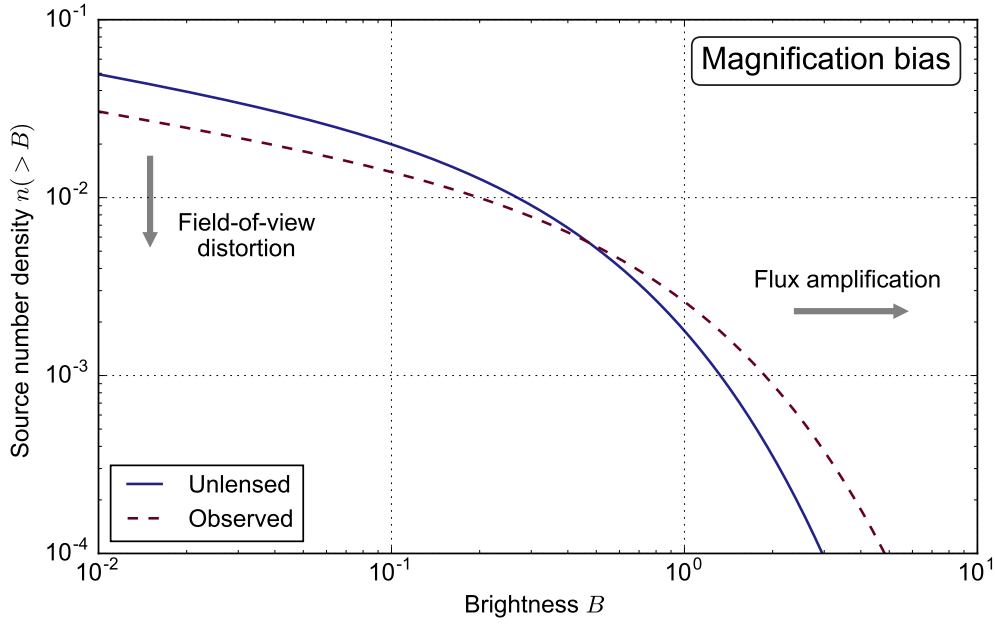
FIGURE 4.9: Illustration of the magnification bias. Weak lensing creates two effects on source number counts: the flux amplification and the field-of-view distortion. The combination of both results in a tilt of the density distribution. The number of luminous galaxies enhances and the number of faint ones depletes. The effect is exaggerated.

4.3.3 Second-order statistics

Lensing scientists (or "lensers") have been interested in different statistical quantities to extract cosmological information from the large-scale structures (LSS). The most studied among these is the second-order statistics, including the power spectrum. Following Schneider (2005), by introducing the Fourier transform of κ:

$$\tilde{\kappa}(\boldsymbol{\ell}) = \int \mathrm{d}^2\boldsymbol{\theta} \; \mathrm{e}^{-\mathrm{i}\boldsymbol{\ell\theta}}, \tag{4.92}$$

the 2D convergence power spectrum $P_\kappa(\ell)$ can be defined as

$$\langle \tilde{\kappa}(\boldsymbol{\ell})\tilde{\kappa}^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta(\boldsymbol{\ell} - \boldsymbol{\ell}') P_\kappa(\ell), \tag{4.93}$$

where δ is the Dirac delta function. Using Limber's approximation (Limber 1953), $P_\kappa(\ell)$ can be related to the matter power spectrum $P_\delta(\ell)$ by (Kaiser 1992)

$$P_\kappa(\ell) = \frac{9H_0^4\Omega_\mathrm{m}^2}{4\mathrm{c}^4} \int_0^{w_\mathrm{max}} \mathrm{d}w \; \frac{g^2(w)}{a^2(w)} P_\delta\left(k = \frac{\ell}{f_K(w)}, w\right), \tag{4.94}$$

where $g(w)$ is the lens efficiency given by Eq. (4.38), and the projection integral is carried out over the comoving distance w up to the limiting value $w_\mathrm{max} = w(a = 0)$. Similarly, the shear power spectrum leads to

$$\langle \tilde{\gamma}(\boldsymbol{\ell})\tilde{\gamma}^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta(\boldsymbol{\ell} - \boldsymbol{\ell}') P_\gamma(\ell). \tag{4.95}$$

Actually, $|\tilde{\gamma}| = |\tilde{\kappa}|$ (see further, Eq. 4.108), so $P_\gamma(\ell) = P_\kappa(\ell)$. Hence in the following, the focus will be put only on $P_\kappa(\ell)$.
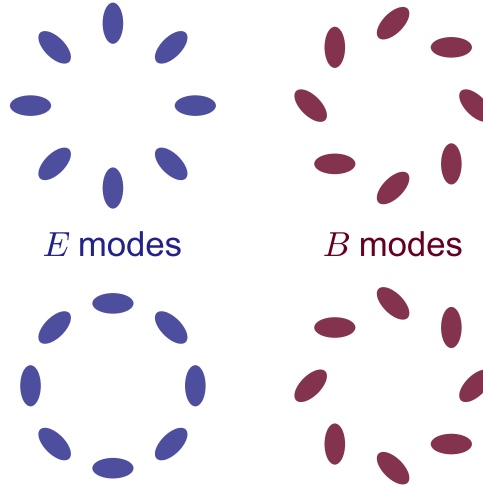
FIGURE 4.10: Illustration of typical patterns from weak lensing E- and B-modes.

Consider now the shear two-point-correlation function (2PCF). Since the shear has two components, its pair product can be separated into three parts, defined as

$$\xi_\pm(\theta) \equiv \langle \gamma_+ \gamma_+ \rangle(\theta) \pm \langle \gamma_\times \gamma_\times \rangle(\theta) \ \ \text{and} \ \ \xi_\times(\theta) \equiv \langle \gamma_+ \gamma_\times \rangle(\theta). \tag{4.96}$$

These correlation functions should be invariant under parity transformation. Under the transformation, $\gamma_+ \to \gamma_+$, $\gamma_\times \to -\gamma_\times$, $\xi_\times$ also changes its sign, so it vanishes and we usually leave it out. How do these quantities relate to $P_\kappa(\ell)$? Following Kaiser (1992), if one transforms γ into $\tilde{\gamma}$, the correlation functions become

$$\xi_+(\theta) = \int_0^{+\infty} \frac{\ell \mathrm{d}\ell}{2\pi} J_0(\ell\theta) P_\kappa(\ell) \ \ \text{and} \ \ \xi_-(\theta) = \int_0^{+\infty} \frac{\ell \mathrm{d}\ell}{2\pi} J_4(\ell\theta) P_\kappa(\ell), \tag{4.97}$$

where J_n is the n-th order Bessel function of the first kind. By the orthonormality of the Bessel functions, these integrals are invertible, thus they become

$$P_\kappa(\ell) = 2\pi \int_0^{+\infty} \theta \mathrm{d}\theta \ J_0(\ell\theta) \xi_+(\theta) = 2\pi \int_0^{+\infty} \theta \mathrm{d}\theta \ J_4(\ell\theta) \xi_-(\theta). \tag{4.98}$$

In real life, the equality (4.98) might not be strictly satisfied due to observational noise and systematics. Following the similarity between the lensing shear and the CMB polarization, the shear can be decomposed into two fields such that one of them satisfies Eq. (4.98), called E-modes; and the other violates Eq. (4.98), called B-modes. The characteristic patterns related to E- and B- modes are illustrated in Fig. 4.10. To construct the B-mode shear, the most convenient way might be to introduce the "imaginary" part (in both usual and mathematical meaning) of the convergence, such that $\kappa = \kappa^E + \mathrm{i}\kappa^B$. Crittenden et al. (2002) and Schneider et al. (2002) showed that in this case, the E- and B- mode power spectra can be written as

$$P_E(\ell) = \pi \int_0^{+\infty} \theta \mathrm{d}\theta \ \Big[J_0(\ell\theta) \xi_+(\theta) + J_4(\ell\theta) \xi_-(\theta) \Big], \tag{4.99}$$

$$P_B(\ell) = \pi \int_0^{+\infty} \theta \mathrm{d}\theta \ \Big[J_0(\ell\theta) \xi_+(\theta) - J_4(\ell\theta) \xi_-(\theta) \Big]. \tag{4.100}$$

Unfortunately, Eqs. (4.99) and (4.100) extend to infinity. It is practical to introduce a filter function to reduce the integration range. Some examples can be found in the literature,

including the top-hat filter (Kaiser 1992), the aperture mass (Schneider et al. 1998), the optimized ring statistic (Fu & Kilbinger 2010), and the Complete Orthogonal Sets of E-/B-mode Integrals (COSEBIs, Schneider et al. 2010). Let $W(\theta)$ be the filter acting on the convergence field and $\widetilde{W}(\ell)$ its Fourier transform. Following Crittenden et al. (2002), Schneider et al. (2002), and Schneider & Kilbinger (2007), one can construct

$$T_+(\theta) = \int_0^{+\infty} \ell \mathrm{d}\ell \; J_0(\ell\theta)\widetilde{W}^2(\ell) \quad \text{and} \quad T_-(\theta) = \int_0^{+\infty} \ell \mathrm{d}\ell \; J_4(\ell\theta)\widetilde{W}^2(\ell), \tag{4.101}$$

which are implicitly linked together by a relation similar to Eq. (4.98). If T_+ vanishes outside an interval $[\theta_{\min}, \theta_{\max}]$ and satisfies the following condition:

$$\int_{\theta_{\min}}^{\theta_{\max}} \theta\mathrm{d}\theta \; T_+(\theta) = 0 = \int_{\theta_{\min}}^{\theta_{\max}} \theta^3\mathrm{d}\theta \; T_+(\theta), \tag{4.102}$$

then T_- also has non-zero values only on $[\theta_{\min}, \theta_{\max}]$, and E- and B-mode information (denoted simply as E and B) can be separated with finite integrations as

$$E = \int_{\theta_{\min}}^{\theta_{\max}} \theta\mathrm{d}\theta \; \Big[T_+(\theta)\xi_+(\theta) + T_-(\theta)\xi_-(\theta)\Big], \tag{4.103}$$

$$B = \int_{\theta_{\min}}^{\theta_{\max}} \theta\mathrm{d}\theta \; \Big[T_+(\theta)\xi_+(\theta) - T_-(\theta)\xi_-(\theta)\Big]. \tag{4.104}$$

More generally, the convergence power spectrum can be decomposed into various E_n and B_n by choosing successively $T_{n,\pm}$ with different finite supports. Finally, as detailed in Sect. 4.3.1, the correlations $\xi_\pm$ in the WL regime can be estimated unbiasedly by counting pairs of $\langle\epsilon\epsilon^*\rangle$:

$$\hat{\xi}_+(\theta) = \frac{\sum_{ij} \omega_i\omega_j\Big(\epsilon_+(\boldsymbol{\theta}_i)\epsilon_+(\boldsymbol{\theta}_j) \pm \epsilon_\times(\boldsymbol{\theta}_i)\epsilon_\times(\boldsymbol{\theta}_j)\Big)}{\sum_{ij} \omega_i\omega_j}. \tag{4.105}$$

Note that the sum runs over all pairs (i, j) fulfilling some binning conditions on $\theta = |\boldsymbol{\theta}_i - \boldsymbol{\theta}_j|$.

Recent surveys have successfully provided interesting cosmological constraints from the 2PCF. Kilbinger et al. (2013) processed the data from CFHTLenS and Jee et al. (2013) studied the Deep Lens Survey (DLS). Both results are consistent with the Wilkinson Microwave Anisotropy Probe (WMAP). Some ongoing surveys such as KiDS and DES are also delivering early-stage results with the shear 2PCF (Kuijken et al. 2015; Hildebrandt et al. 2016; The Dark Energy Survey Collaboration et al. 2015).

4.3.4 Higher-order statistics

Higher-order statistics intrigue cosmologists for two reasons. First, as mentioned earlier, there is no simple solution for modelling the nonlinear matter power spectrum, which limits the amount of the information extracted from the 2PCF. Second, the 2PCFs only retain the Gaussian information. Unlike the CMB which can almost be fully described by its Gaussian part, the WL field is rich of non-Gaussianities. Cosmological information encrypted in the non-Gaussian part can be complementary to the power spectrum, and this can be extracted by higher-order statistics, some of which I display in the following.

Bispectrum

A common option of higher-order statistics is the three-point-correlation function (3PCF). As the 2PCF is related to the power spectrum, the 3PCF can be related to the bispectrum, and then to the matter distribution which depends on cosmology (Schneider & Lombardi 2003; Takada & Jain 2003a; Zaldarriaga & Scoccimarro 2003) (also Takada & Jain 2003b,c; Schneider et al. 2005). Similar to the power spectrum, the infinite integral problem also arises from observation. To overcome this obstacle, one can focus on third-order moments which summarize the local 3PCF information (Kilbinger & Schneider 2005). For the data-driven studies, Semboloni et al. (2011) took the data from the Hubble Space Telescope (HST) Cosmological Evolution Survey (COSMOS) and derived the first cosmological constraints with the third-order statistics. Fu et al. (2014) combined the 2PCF and 3PCF, and showed that parameter influence is improved compared to the 2PCF-only studies. For the even higher order, Takada & Jain (2002) addressed the kurtosis of the shear field, and found that its main contribution comes from massive halos with $M > 10^{14} M_\odot$.

Minkowski functionals

Minkowski functionals (MFs) capture the morphology of a considered region by extracting some characteristic quantities. For a given threshold ν and a smoothed 2D field $K(\boldsymbol{\theta})$, MFs are defined as

$$V_0(\nu) \equiv \int_{S_\nu} \mathrm{d}S, \quad V_1(\nu) \equiv \frac{1}{4} \int_{\partial S_\nu} \mathrm{d}\ell, \quad \text{and} \quad V_2(\nu) \equiv \frac{1}{2\pi} \int_{\partial S_\nu} \mathrm{d}\ell \, \mathcal{K}, \tag{4.106}$$

where $S_\nu \equiv \{\boldsymbol{\theta} | K(\boldsymbol{\theta}) > \nu\}$, $\partial S_\nu \equiv \{\boldsymbol{\theta} | K(\boldsymbol{\theta}) = \nu\}$, and $\mathcal{K}$ is the geodesic curvature along the contours. One can see at ease that $V_0(\nu)$ is the area above the threshold, and $V_1(\nu)$ counts for the boundary length. If $K(\boldsymbol{\theta})$ is a smoothed Gaussian random field, MFs can be computed analytically (Sato et al. 2001), whereas Taruya et al. (2002) provided predictions for a log-normal field. Among the recent studies, two series of work focus on application to WL. For CFHTLenS data, forecasts using the Fisher formalism (Shirasaki et al. 2012; Shirasaki & Yoshida 2014) and cosmological constraints (Kratochvil et al. 2012; Petri et al. 2013, 2015) were obtained with MFs.

Peak counts

Recently, *weak-lensing peak counts* have become a popular research topic. Defined as the local maxima (Fig. 4.11) of the aperture mass (see Sect. 8.2.2) or convergence, peaks are tracers of massive structures, particularly galaxy clusters and DM halos. The abundance of these objects is a cosmology-dependent statistic: whereas the 2PCF is modelled via the convergence and matter power spectra, peak counts are modelled via the halo mass function and the collapse theory. Unlike optical richness, the X-ray luminosity or temperature, or the Compton-y parameter from the Sunyaev–Zel'dovich (SZ) effect, the connection of WL peak counts to the mass function is more direct since the mass-to-shear relation is simpler than the mass-to-light relation, which depends on baryonic matter (Schneider 1996). Besides, peaks are straightforward to identify in the surveys. These reasons make WL peak counts a potential cosmological probe.
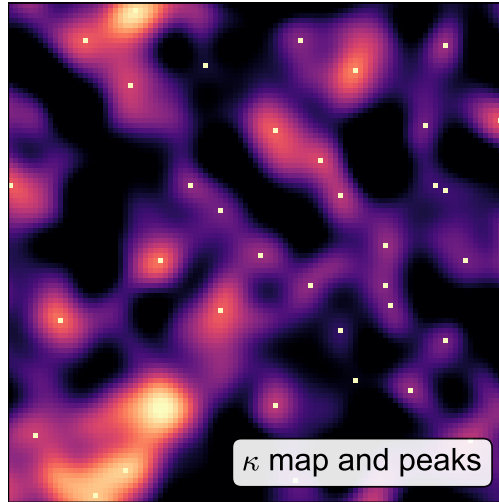
FIGURE 4.11: Peaks from a weak-lensing map.

4.4   Observational and modelling challenges

After introducing the basis of WL and its measurement, I would like to turn to the practical side and discuss the observational challenges that cosmologists might encounter for WL surveys. WL is mainly observed in the optical and near-infrared bands, since source counts are highest in these wavelengths. The "detector" are cameras with photometers and spectrometers to estimate redshifts. Images need to be processed in order to separate individual galaxies and to determine the lensing signal.

Shape measurement

Measuring the galaxy shape up to a very high precision is important since WL surveys aim to measure very small shape distortions. For example, Euclid, a stage IV WL survey, aims to measure the ellipticity with an error $\leq 2 \times 10^{-4}$ (Laureijs et al. 2011). However, such a task is difficult. One main reason is the point spread function (hereafter PSF). The PSF characterizes the fact that a distant point-like source always results in a blurred image in an observation. The origin of the blurring could be, for example, the optical system of the telescope, and the perturbations in Earth's atmosphere for a ground-based survey.

In general, the PSF varies with position, wavelength, and time. A straightforward way to handle it is to consider the images of stars as some local PSF samples since stars are so small that they can be considered as point sources, and to interpolate over the field of view. Various sophisticated methods using Bayesian model fitting (Miller et al. 2007) and super-resolution (Ngolè Mboula et al. 2015) also exist.

Besides the PSF, the pixelization adds some additional uncertainty to shape estimation. Even though the resolution of ground-based telescopes reaches the order of sub-arcseconds, the size of distant faint galaxies is not much larger. The instrumental and photon noise adds another dimension to the complex shape measurement problem. Today, robust methods are strongly required as the data volume increases to derive precise cosmological constraints.

FIGURE 4.12: Left panel: a beautiful high-resolution image of NGC 201. Right panel: the same image blurred with a Gaussian PSF. (Source: ESA/Hubble)

Shape noise

The term *shape noise* qualifies the randomness of galaxy intrinsic orientation. In some studies, one can read "shot noise" instead. This terminology comes from the consideration that the expected noise for the convergence is proportional to $1\sqrt{N_{\mathrm{gal}}}$ where N_{gal} is the number of averaged galaxies. However, this is misleading since (1) shape noise does not follow the Poisson law (Poisson 1837), and (2) the proportionality to $1\sqrt{N_{\mathrm{gal}}}$ is no longer valid if the convergence is obtained with a weighted average. Therefore, the term "shot noise" should be avoided.

Theoretically, the distribution of the shape noise is unknown. However, since lensing is in general weak, $\epsilon \approx \epsilon^{(\mathrm{s})} + g$ and the overall distribution on $\epsilon^{(\mathrm{s})}$ will be similar to the one of ϵ. In practice, cosmologists assume a (truncated) normal distribution of the same variance $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2$ for both components. Throughout this work, I define the variance of the *intrinsic ellipticity dispersion* σ_ϵ^2 as the sum of variances of both components, i.e. $\sigma_\epsilon^2 \equiv \sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2$. Alternatively, some studies define σ_ϵ^2 as the variance of one component, which results in a difference of factor $\sqrt{2}$. The value of σ_ϵ could potentially depend on cosmology on the galaxy evolution level. However, cosmologists usually account for it as an external parameter.

From CFHTLenS, the dispersion is $\sigma_\epsilon \approx 0.4$ (Kilbinger et al. 2013). From N-body simulations, only very few galaxies carry a convergence larger than 0.1. As a result, we are facing a problem where the signal is totally submerged in the noise. Therefore, filtering becomes an important step. I will discuss this topic in Chap. 8.

Mask effects

Like all cosmological surveys, real-life WL data are partially masked. The origin of these masks is various. A very common type of mask is bright stars. Light from these objects diffracts in the telescope and creates spike-like artifacts. Normally, these spikes can be modelled by the PSF. However, if they are too large and too luminous, nearby images will be contaminated. A usual way to handle these is to cut out a disk-like region centered at the star position (see
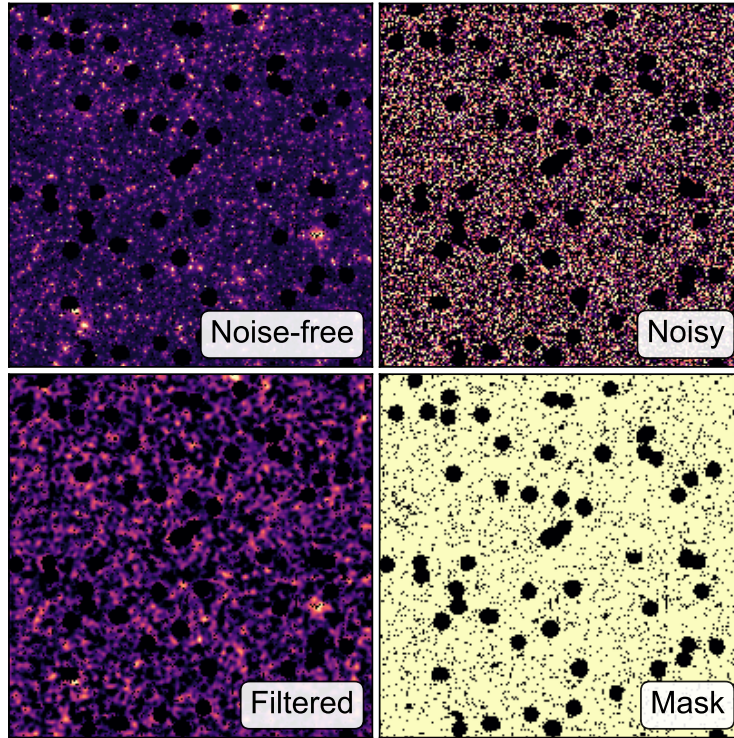
FIGURE 4.13: Example of lensing maps with a mask taken from CFHTLenS data.

Fig. 4.13).

Also, the charge-coupled devices (CCD) of the camera do not perform perfectly. Individual abnormal pixels can create dotted masks, and depending on the geometry of electronic components, these masks can also be striped.

Finally, because WL probes in optical and infrared bands, the Milky Way becomes naturally the greatest mask for any full sky survey. The star and gas emission in the galactic plane makes approximately between one-third and one-half of the sky unavailable for the survey. Also, large foreground galaxies which do not contribute to lensing signals contaminate astronomical images. An optimistic size of the full-sky survey contains actually $\lesssim 25000 \ \mathrm{deg}^2$.

Inversion problem and mass-sheet degeneracy

Even though we saw earlier that the magnification helps to estimate the convergence, this is intractable without stacking due to a low signal-to-noise ratio. Because of this reason, a κ map should be determined from a γ map (more exactly a g map). As we have seen, κ and γ are related via Eq. (4.29). A direct way to derive κ is to "invert" this relation. Actually, if we go to Fourier space, derivatives become multiplications by wave numbers. Thus, Eq. (4.29) leads to

$$\hat{\kappa} = \frac{\ell_1^2 + \ell_2^2}{2}\hat{\psi}, \quad \hat{\gamma}_1 = \frac{\ell_1^2 - \ell_2^2}{2}\hat{\psi}, \quad \hat{\gamma}_2 = \ell_1 \ell_2 \hat{\psi}. \tag{4.107}$$

Readers can see that both $\hat{\gamma}_1$ and $\hat{\gamma}_2$ can be used to express $\hat{\kappa}$. To minimize the noise variance, we can set $\hat{\kappa} = (1-\alpha)\hat{\gamma}_1 + \alpha\hat{\gamma}_2$ and derive it with regard to α to find the optimum. The same

result can be recovered by treating $\hat{\gamma}$ directly as $\hat{\gamma}_1 + \mathrm{i}\hat{\gamma}_2$:

$$\hat{\gamma} = \frac{(\ell_1 + \mathrm{i}\ell_2)^2}{2}\hat{\psi}, \quad \text{so} \quad \hat{\kappa} = \frac{\ell_1^2 + \ell_2^2}{(\ell_1 + \mathrm{i}\ell_2)^2}\hat{\gamma}. \tag{4.108}$$

The final results from both methods, called Kaiser-Squires inversion (Kaiser & Squires 1993), are respectively

$$\hat{\kappa} = \frac{\ell_1^2 - \ell_2^2}{\ell_1^2 + \ell_2^2}\hat{\gamma}_1 + \frac{2\ell_1\ell_2}{\ell_1^2 + \ell_2^2}\hat{\gamma}_2 \quad \text{and} \quad \hat{\kappa} = \frac{\ell_1^2 - \ell_2^2 - 2\mathrm{i}\ell_1\ell_2}{\ell_1^2 + \ell_2^2}\hat{\gamma}, \tag{4.109}$$

which are actually identical by simple algebra and Eq. (4.107).

However in reality, only g is directly measurable. If we use the weak-lensing approximation, i.e. $\gamma = (1 - \kappa)g \approx g$, then κ will be biased high. To correct this bias, one can iteratively apply Eq. (4.109), i.e. using the convergence from the previous iteration $\kappa^{(t-1)}$ to adjust the shear $\gamma^{(t)} = g/(1 - \kappa^{(t-1)})$ and to yield a new estimation of the convergence $\kappa^{(t)}$ from Eq. (4.109). An alternative method using a similar technique is proposed by Seitz & Schneider (1995).

Note that Eq. (4.109) does not possess any solution when $\ell_1 = \ell_2 = 0$. In real space, this wave mode corresponds to the constant term, say κ_0. It means that the global level of the convergence map is undetermined by this method. In principle, this degeneracy should not exist if we have the following idealistic lensing information: if κ is known over the full sky at a fixed source depth w, then the mean of κ, which is κ_0, is necessarily 0. The reason is that Eq. (4.36) does not integrates over the mass density but the density contrast δ, and the mean of the contrast over the sky is zero In reality, cosmologists possess only finite lensing samples which are located in the non-masked field of view and are not equally distant. These facts cause $\kappa_0 \neq 0$. However, if the field is large $\kappa_0 \approx 0$ is a good approximation.

A related concept is proposed by Falco et al. (1985). Consider, for a source catalogue and a gravitational lens, the following change of variables:

$$1 - \kappa'(\boldsymbol{\theta}) = \lambda(1 - \kappa(\boldsymbol{\theta})) \quad \text{or} \quad \kappa'(\boldsymbol{\theta}) = \lambda\kappa(\boldsymbol{\theta}) + (1 - \lambda), \tag{4.110}$$

where λ is an arbitrary constant and $\kappa(\boldsymbol{\theta})$ is the true convergence from this system. In the cluster lensing framework, this λ-transformation implies $\Sigma'(\boldsymbol{\theta}) = \lambda\Sigma(\boldsymbol{\theta}) + (1 - \lambda)\Sigma_{\mathrm{crit}}$ from Eq. (4.45). Thus, if the profile is circular, from Eq. (4.56) we obtain $\gamma'_+(\boldsymbol{\theta}) = \lambda\gamma_+(\boldsymbol{\theta})$. More generally, any distribution of Σ can be considered as the convolution by a properly chosen kernel of a punctual mass (which is a "circular" profile), so $\gamma'(\boldsymbol{\theta}) = \lambda\gamma(\boldsymbol{\theta})$ everywhere. Therefore, the reduced shear is invariant and the transformation (4.110) leads to a series of admissible solution $(\kappa(\boldsymbol{\theta}), \gamma(\boldsymbol{\theta}))$ for the same observation. This is called the *mass-sheet degeneracy*. For example, an Einstein ring could be an image of a strongly lensed background galaxy, or a lensing-free image of a huge ring-like structure.

Sometimes, the mass-sheet degeneracy also refers to κ_0, claiming that κ_0 and λ are the same degree of freedom. This is true when the field of view is small. Actually, λ is a local degeneracy since the λ-transformation is only applied to a circular area neighboring the cluster (Falco et al. 1985), whereas κ_0 is a global degeneracy which accounts for the whole field of view. Fundamentally, κ_0 and λ are still different concepts.

Source redshifts

The lensing signal depends not only on the distance from the observer to the lens, but also on the one to the source. However, because of the broad kernel which acts in Eq. (4.36), a
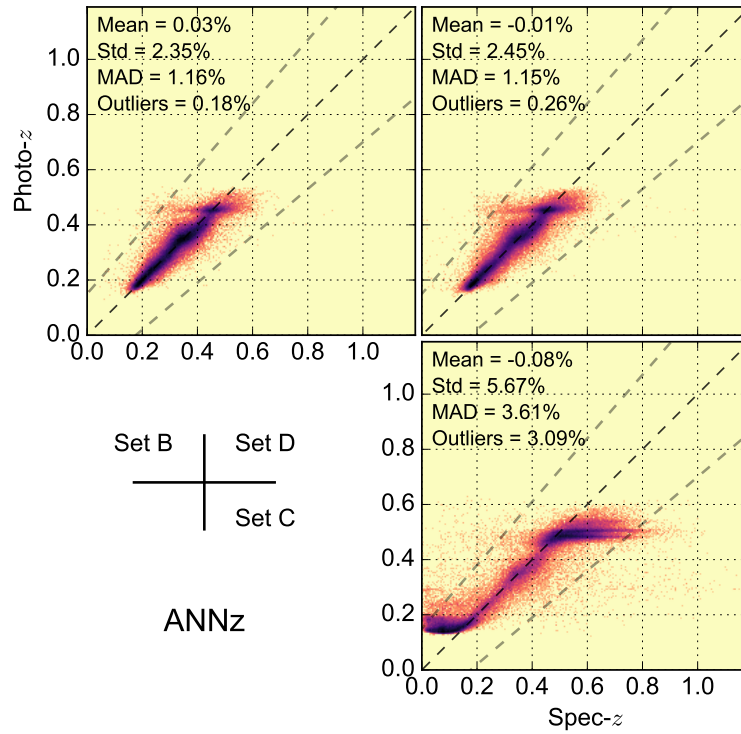
FIGURE 4.14: Comparison between the spectroscopic and photometric redshifts for three data sets constructed from Sloan Digital Sky Survey (SDSS) data releases 12. All three panels show the scatter of the photometric redshift estimation from samples with the spectroscopic redshift measurement (considered as the truth). In the bottom right panel, we can see some catastrophic failures, where the estimation error is flagrant.

very high precision on source redshifts is not necessarily required. Therefore, most surveys count on the photometric technique (Connolly et al. 1995; Benítez 2000) to determine the source redshifts for a wide field, which is less precise but efficient, while the spectroscopic technique is only applied at a relatively restricted field size, to determine redshifts with high precision and to calibrate the photometric results. A very straightforward way to include the source information is to transform Eq. (4.36) into an integral over the redshift distribution, reasoning in a probabilistic way. An alternative is to split the source catalogue into different slices depending on their depth, and to perform a tomographic analysis. This is particularly interesting to study cosmological evolution for different epochs, which helps breaking some parameter degeneracies.

In this context, the focus of redshift studies for WL has been put on modelling the errors on photometric redshifts (photo-z) and understanding how these errors propagate to the lensing signal. An inevitable phenomenon for photometric redshifts are the "catastrophic" failures: the difference between the true and estimated values to be several times larger than the characteristic error. These failures are problematic if the proportion is large.

Recently, a technique taking advantage of the galaxy correlation function to derive the redshift distribution has been rediscovered and promoted (Seldner & Peebles 1979; Roberts & Odell 1979; Newman 2008; Schmidt et al. 2013). Such a technique has already been used by Hildebrandt et al. (2016) as an alternative to photo-z.

Random orientation

Intrinsic alignment
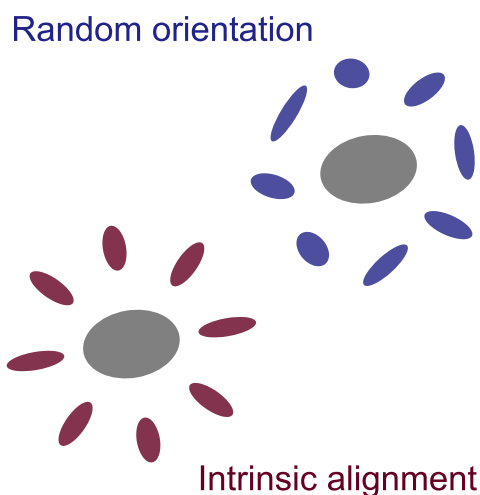
FIGURE 4.15: Illustration of intrinsic alignment. Galaxies can be influenced by the tidal force of nearby halos and aligned with the radial direction. The effect is exagerated in the figure.

Intrinsic alignment

As we have seen in Sect. 4.3.1, the interpretation of the lensing signal and convergence maps is based on the isotropy hypothesis of galaxy orientation. The intrinsic galaxy orientation should not possess any privileged direction. This is however not true in reality. First, galaxies that are close to a halo are subject to a tidal force, and are constantly stretched in radial direction during their formation. Second, the angular momenta of the halo and the galaxy can be correlated. These physical processes break the isotropy hypothesis locally, which causes a phenomenon called *intrinsic alignment* (IA). If the isotropy hypothesis is still applied, the presence of IA creates spurious lensing signals (Heavens et al. 2000; Bridle & Abdalla 2007). In terms of second-order statistics, this contamination can be separated into two terms: an intrinsic-intrinsic correlation and a shear-intrinsic correlation. Removing these contaminations is challenging. Alternatively, some physically motivated models for the intrinsic shape of galaxies have been proposed (Hirata & Seljak 2004; Bridle & King 2007). This provides a solution for modelling using a forward approach. In any case, studies have already shown that neglecting or mis-modelling IA could introduce important biases on WL signals (e.g. Kirk et al. 2015).

Baryonic effects

At small scales, structure formation is influenced by baryonic physics. To solve the *overcooling problem*, i.e. the star formation rate is higher than expected, the feedback from active galactic nuclei (AGN) are required during galaxy formation (Dubois et al. 2013; Okamoto et al. 2014). Recent hydrodynamic simulations showed that including this feedback could change the DM halo profiles (Duffy et al. 2010; Martizzi et al. 2012), which changes then the lensing signal at small scales. Most WL studies until recently, especially those which focus on second-order statistics, do not account for baryonic effects since at large scales this is insignificant. However, if cosmologists desire to probe the power spectrum at high-ℓ or to study peak counts, including baryons would be indispensable.
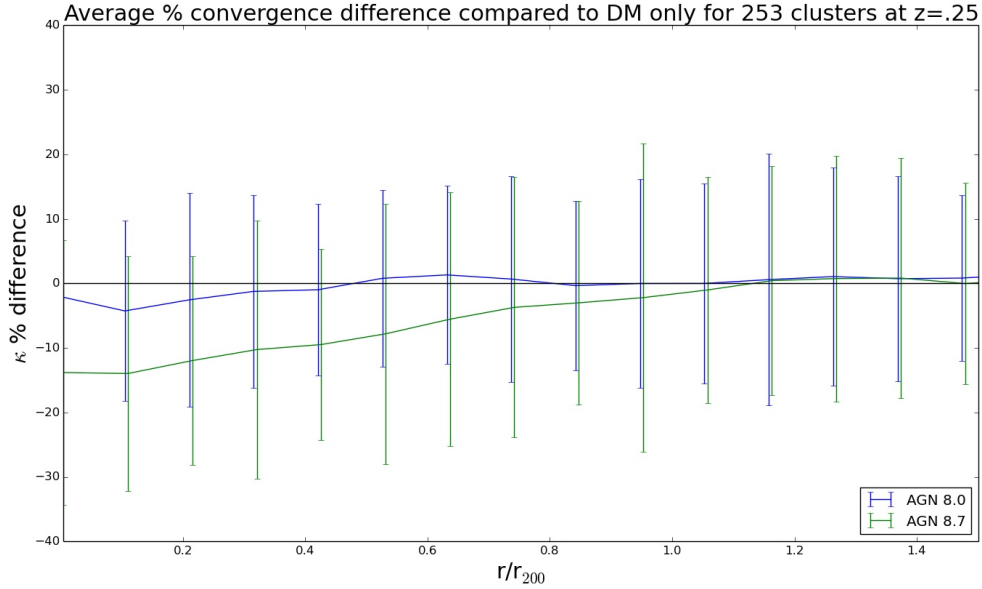
FIGURE 4.16: Preliminary result of the difference between profiles of DM-only clusters and those with baryonic processes. In the extreme AGN case (green), baryons can change the lensing signal by 10–15% near the center of the clusters. (Source: Brandyn Lee)

Nonlinear spectrum

There exist several methods to extract cosmological information from WL. The one which acquired the greatest focus to date is the power spectrum. This is mostly motivated by the fact that the matter spectrum can be well modelled by theory on large scales. However on small scales, complex gravitational interactions make the matter distribution nonlinear and non-Gaussian, and invalidate the prediction from the linear theory. Until now, two families of corrections have been proposed, which are perturbation theory (e.g. Makino et al. 1992; Bernardeau et al. 2002) and effective field theory (e.g. Baumann et al. 2012; Carrasco et al. 2012). Nevertheless, neither of them can provide a satisfactory result for scales below few Mpc/h (Cooray & Sheth 2002; Smith et al. 2003; Carrasco et al. 2012). For this reason, modelling using N-body simulations is an interesting option. This difficulty also provides a great motivation to explore WL observables other than the power spectrum.

4.5 Weak-lensing peak counts: state of the art

A bit of history

The idea of counting peaks in WL was first proposed by Schneider (1996). The original argument was that at signal-to-noise ratio (S/N) of $\gtrsim 5$, peaks are presumably real halos, so we probe directly the high-mass end of the mass function to discriminate cosmologies. The detection was expected to be carried out by the aperture mass (Kaiser et al. 1994; Schneider 1996). The following studies of the same series also stayed in this "lens-count" approach: Kruse & Schneider (1999, 2000) estimated the expected number of detections for different cosmological models, and provided an exponential model to describe the fast decreasing tail;
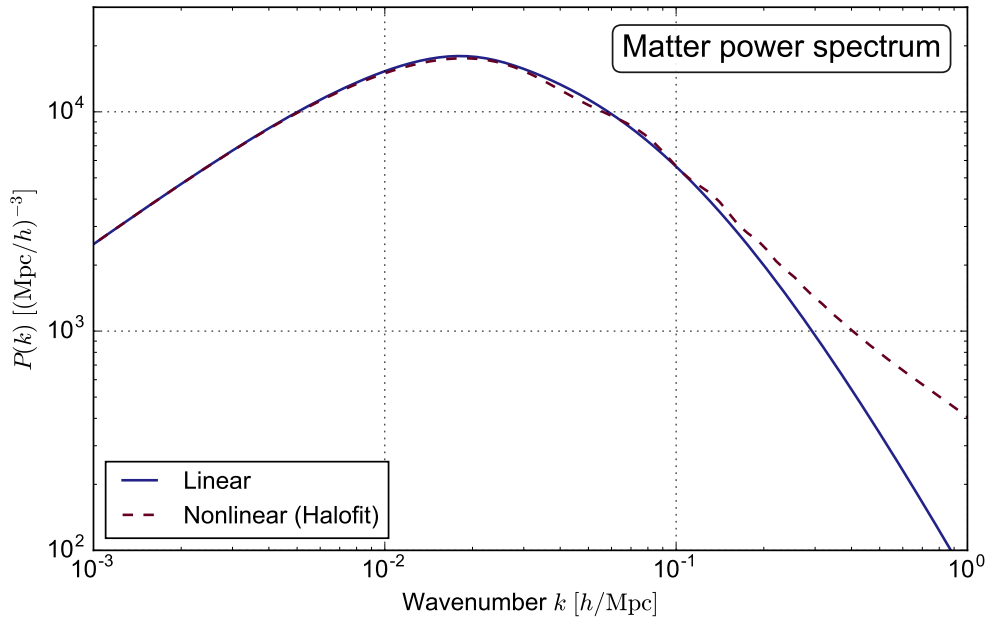
FIGURE 4.17: Comparison between the linear and the nonlinear power spectrum. The nonlinear spectrum is computed using the HALOFIT algorithm. We can see that at small scales the linear spectrum under estimates the power. Also, the spectrum oscillates for $k > 10^{-2}$. This is due to the baryonic acoustic oscillations (BAO).

Reblinsky et al. (1999) validated the previous works by N-body simulations. On the other hand, Jain et al. (2000), Munshi & Jain (2000), and Bernardeau & Valageas (2000) studied the full-κ distribution (this is actually one of the Minkowski functionals!) modelled with appropriate noise properties and derived its cosmological dependency.

Later, Van Waerbeke (2000) and Jain & Van Waerbeke (2000) combined both concepts and modelled directly the peak function without distinguishing between true and spurious peaks. Since then, studies of WL peak counts can be roughly classified into two categories (Lin et al. 2016). The first category is concerned with *cluster-oriented purposes* which stays close to the original philosophy of Schneider (1996) to establish explicitly the *selection function*, i.e. the mass-convergence correspondence. These studies are usually interested in high S/N, positional offsets, peak-height variations, and projection effects focusing on indicators such as purity and completeness. They may also cross-check with optical richness and X-ray data. The other category focuses on *cosmology-oriented purposes* and adopts the idea of Jain & Van Waerbeke (2000). These studies make the selection function implicit. They attempt to directly model WL peak counts from cosmological models, including contributions from massive clusters, projections of LSS, spurious signals, or a mixture of all of these cases [♯2]. This is precisely the focus of this thesis.

In the following, I outline some remarkable series of works on WL peak counts, organized by their respective working group. Before this, it should be appropriate to mention first Bartelmann et al. (2001, 2002) who proposed to model the expected detection as the sum of the theoretical prediction and a Gaussian fluctuation, and who argued that cluster counts could be used to constrain the equation of state of the dark energy. These works influenced some of the following results.

[♯2]Of course, this is not a strict classification, since cluster detection and peak counts rely on the same peak-finding technique. Some work satisfy both purposes.

Tokyo

One of the series is led by Hamana et al. (2004). From N-body simulations, the authors studied the mass-convergence correspondence in detail, and extended the analytical cluster-count prediction model from Kruse & Schneider (2000) and Bartelmann et al. (2002) by emperically establishing the selection function. This model was improved by Hamana et al. (2012) as halo triaxiality has been taken into account. Modelling the lens geometry fluctuation, the projection from LSS, and the shape noise, Hamana et al. (2015) applied the model from previous works to the Subaru Suprime-Cam data and obtained cosmological constraints. Recently, the focus of this group has been put on the ongoing HSC survey. Osato et al. (2015) drew particular attention on the baryonic impact on the WL power spectrum and peaks, providing forecast for HSC. Shirasaki et al. (2016a) and Higuchi & Shirasaki (2016) examined the sensitivity of cluster counts to $f(R)$ parameters by Fisher information (Fisher 1922) and N-body simulations, respectively validating the feasibility of the probe. Later, Shirasaki (2016) tried to model peak counts analytically by combining different orders of moments using local-Gaussianized transformation. Finally, Shirasaki et al. (2016b) investigated Fisher information provided by different WL observables, including peak counts.

Heidelberg

Another remarkable series is led by Maturi et al. (2005) focusing on optimizing cluster detection. Maturi et al. (2005) proposed a filter optimized for identifying NFW-like halos. Note that Schirmer (2004 [♯3], see also Hetterscheidt et al. 2005) also proposed an optimized filter for the same issue. The difference is that Schirmer (2004) fitted an analytical function to the expected signal from an NFW profile, while Maturi et al. (2005) argued that in order to detect clusters better, the projection from LSS should be treated as noise and taken into account. The author proposed then to consider the new noise variance from the sum of the LSS spectrum and the galaxy shape noise, and constructed a filter adapted to the NFW signal and the total noise. Pace et al. (2007) compared this filter with other commonly used ones. Then, Maturi et al. (2007) applied this method to the GaBoDS data in parallel with Schirmer et al. (2007).

Later, Maturi et al. (2010) proposed an analytical model to model peak counts. While Bartelmann et al. (2002) and Hamana et al. (2004) model the selection function either by a Gaussian fluctuation or N-body simulations, this new approach, more aligned with cosmology-oriented purposes, is directly based on the peak theory of the Gaussian random field (Bardeen et al. 1986; Bond & Efstathiou 1987). The definition of peaks has been changed, from the local maxima, pixels with values higher than their eight neighbors, to the *contiguous areas* with values above a given threshold. The motivation is that the new definition simplifies the analytical expression. After that, Maturi et al. (2011) extended the previous work and yielded the Fisher forecast on f_{NL}. Recently, Reischke et al. (2016) came out with a correction on very high S/N for the model of Maturi et al. (2010).

Philadelphia-Bonn

In this working group, Marian & Bernstein (2006) started with forecasting the constraining power of cluster counts. Then, using N-body simulations, Marian et al. (2009, 2010) examined

---

[♯3]PhD thesis not available on NASA/ADS, but via `http://hss.ulb.uni-bonn.de/2004/0326/0326.htm`.

the projection effect in the noise-free case, and Marian et al. (2011) provided a quick look on sensitivity of peaks on probing $f_{\rm NL}$. Many Fisher analyses have been performed by the authors. They focused on the mass function (Smith & Marian 2011), $w_0^{\rm de}$ and ΛCDM parameters (Marian et al. 2012), $f_{\rm NL}$ with 2PCF and peak counts (Hilbert et al. 2012), and $w_0^{\rm de}$ and ΛCDM parameters with peak abundance combined with γ_+ profiles and the peak-peak correlation (Marian et al. 2013). In particular, Marian et al. (2012) also proposed a hierarchical filtering algorithm for mass reconstruction. This actually coincides with the idea of the matched filter. As a remark, the authors of these studies claimed to aim for cosmology-oriented purposes [♯4]. However, the methods and tools turned out to stay close to cluster-oriented approaches.

Garching-Bonn-DES

Dietrich & Hartlap (2010) provided the first realistic forecast on the $\Omega_{\rm m}$-σ_8 constraints. Modelling with a large amount of N-body simulations, they explored the joint constraining power of peak counts and the 2PCF, for obvious cosmology-oriented purposes. They showed that peaks alone constrain $\Omega_{\rm m}$ and σ_8 better than the 2PCF. Dietrich et al. (2012) studied the impact of the LSS projection effect and the noise on peak positions. Later, Kacprzak et al. (2016) followed Dietrich & Hartlap (2010) and performed the $\Omega_{\rm m}$-σ_8 constraints with the DES data.

New York

At the beginning, this series of studies followed a pure cluster-oriented approach. Wang et al. (2004) performed Fisher analyses with X-ray, SZ, and WL-selected clusters, while Fang & Haiman (2007) considered a similar problem with WL clusters and 2PCF. However, the strategy turned into a cosmology-oriented focus later. Wang et al. (2009) investigated the properties of the full κ PDF. As mentioned earlier, the κ PDF is one of the MF. Since then the group developed a remarkable series of studies on both peak counts and MF. Similar to Dietrich & Hartlap (2010), they modelled the observables with N-body simulations. First, Kratochvil et al. (2010) tested the sensitivity to cosmology of peak counts. Then, the authors focused on three parameters: $\Omega_{\rm m}$, σ_8, and $w_0^{\rm de}$. Yang et al. (2011) examined the number of halos associated to a peak and yielded Fisher forecasts. In this study they found a counterintuitive fact: the noise enhances the peak S/N. Later, Yang et al. (2013) investigated the baryonic impact on the WL peaks and power spectrum. They discovered that low peaks with S/N between 1 and 3.5 not only are robust to baryonic effects, but also contain valuable cosmological information. After studying the magnification bias, Liu J. et al. (2014) concluded that this is negligeable for current small-coverage surveys such as CFHTLenS, but needs to be taken into account for Euclid and LSST. Meanwhile, MF are studied by Kratochvil et al. (2012), Petri et al. (2013), and Petri et al. (2015). Liu J. et al. (2015a) applied the previous works to the CFHTLenS data and obtained cosmological constraints, and Liu & Haiman (2016) examined the peak-peak correlation from the same data set. Finally, Zorrilla Matilla et al. (2016) compared the fast modelling method proposed by Lin & Kilbinger (2015a) with N-body simulations and examined its systematics in detail. The authors found a good agreement and proposed several potential improvements. Note that

---

[♯4]For example, we can read from Marian et al. (2009): "However, our goal here is not to establish the correspondence between the two-dimensional and three-dimensional masses of individual halos, but the cosmology dependence of the shear-peak abundance."

FIGURE 4.18: Saclay and its charming fields. (Are they Gaussian?)

this group is also involved in the preparation for LSST, where Bard et al. (2013) dealt with the impact of measurement errors on peak counts, and Bard et al. (2016) studied masks.

Beijing

Like many other working groups, this series of studies also appeared to be interested in cluster counts first, and then turned to a cosmology-oriented method afterwards. The series started with Tang & Fan (2005) who tried to establish a formal selection function for NFW lensing peaks. Then, Fan (2007) attempted to study the impact from IA on WL peaks. The author modelled IA as an enhancement of the noise variance, thus proposed a correction to cluster counts. After that, Fan et al. (2010) chose to use an analytical prescription to model peak counts based on the Gaussian peak theory. In the end, this approach is similar to Maturi et al. (2010) except for two differences. The first one is the definition of peaks for which Fan et al. (2010) opted for the local maximum. The second one is the filter shape, chosen to be Gaussian. However, since the filter from Maturi et al. (2010) and the Gaussian kernel can be exchanged with each other and be implemented in both models without difficulty, this difference is not crucial. Jiao et al. (2011) compared the Gaussian kernel with the MRLens nonlinear filter (see Starck et al. 2006, Sect. 8.3.3) on information extraction. Afterwards, Liu X. et al. (2014) studied the mask effect. They proposed a correction for the model of Fan et al. (2010) to handle the bias generated by missing data. Recently, using the Canada-France-Hawaii Telescope (CFHT) Stripe 82 data, Shan et al. (2014) gave the observed peak function and compared with the model, and Liu X. et al. (2015b) derived the cosmological constraints on Ω_{m}, σ_8, and halo concentration parameters. Finally, Liu X. et al. (2016) constrain parameters from the $f(R)$ theory using the CFHTLenS data.

Saclay

In this series, Pires et al. (2009a) compared different non-Gaussian observables by using different filters for convergence reconstruction, including the MRLens nonlinear method. They concluded that peak counts outperformed other higher-order statistics for discriminating

between cosmological models. A similar study is carried out by Pires et al. (2012), where they investigated the optimal scale for cosmological information extraction. Meanwhile, Bergé et al. (2010) addressed a similar question with a different approach. The author performed the Fisher analysis with the power spectrum, bispectrum, and halo counts to explore the constraining power of these quantities.

Then, for full cosmology-oriented purposes, Lin & Kilbinger (2015a, hereafter Paper I) proposed a stochastic approach to model WL peak counts (see Sect. 5.2). The model adopts the halo approach, and in a forward way computes the WL peak counts from the considered mass function. This approach has the potential to include complex survey conditions in a straightforward way. The authors showed that the new model bypasses the costly N-body computation without losing consistency. Next, Lin & Kilbinger (2015a, hereafter Paper II) were interested in parameter constraint strategies. They discovered that the Gaussian likelihood is a good approximation. However, if the covariance is assumed to be independent from cosmology, the constraining power might be underestimated. The authors also showed that by using approximate Bayesian computation (ABC, see Chap. 7), cosmological constraints can be obtained without a likelihood function in a much faster way. Also, Lin et al. (2016, hereafter Paper III) explored different filtering techniques, finding that using multiscale information from compensated filters extract more cosmological information from peak counts than the Gaussian function.

Some other studies

Further studies are presented below. White et al. (2002) investigated purity and completeness of cluster detection and claimed that noise and biases on mass were too large for cluster counts. Meanwhile, Weinberg & Kamionkowski (2002, 2003) investigated the cluster abundance and analyzed its detection by WL for different dark-energy models. Hennawi & Spergel (2005) proposed a tomographic matched filter adapted for cluster detection at different scales and redshift. This work involves comparisons between different filter sizes, filter shapes, and tomography binwidths, using purity and completeness as indicators. Jaroszyński & Kostrzewa (2010) studied the mass-convergence correspondence in the noiseless case. Li (2011) used wavelets to filter the convergence maps and studied the sensitivity of peaks to w_0^{de}. The impact from the halo concentration has been addressed by King & Mead (2011), while Schmidt & Rozo (2011) compared different filters mentioned in the literature. VanderPlas et al. (2012) applied inpainting to the convergence maps and discovered that spurious peaks were reduced by a factor of 3. Romano et al. (2012) provided a simple cluster-count forecast for the Euclid mission.

Some works yielded constraints from data. Dahle (2006) became the first to reconstruct the cluster mass function with WL and to constrain Σ_8 from it. This study was also extended to the constraint on the neutrino mass by combining the WL mass function with CMB, supernovae of type Ia (SNIa), and galaxy clustering (Kristiansen et al. 2007).

Concerning Fisher analysis, readers find estimates on Ω_{m}, σ_8, w_0^{de} and w_a^{de} (Shapiro & Dodelson 2007), similar parameters with tomography (Takada & Bridle 2007), a joint analysis with halo abundance, primordial non-Gaussianity (Dalal et al. 2008), another joint analysis with galaxy and halo clustering (Yoo & Seljak 2012), estimates on $f(R)$ parameters (Cardone et al. 2013), on M-c relation parameters (Mainini & Romano 2014; Cardone et al. 2015), and ΛCDM parameters using tomography (Martinet et al. 2015).

Also, WL-selected clusters have been investigated in several surveys, e.g. from the Subaru

Suprime-Cam data (Miyazaki et al. 2002), the first Focal Reducer Spectrograph (FORS1) field of the Very Large Telescope (VLT) (Hetterscheidt et al. 2005), the DLS data (Wittman et al. 2006), the CFHT Legacy Survey Deep field (CFHTLS-Deep) (Gavazzi & Soucail 2007), the Bonn lensing-optical-X-ray catalogue (Dietrich et al. 2007), the Garching-Bonn Deep Survey (GaBoDS) (Schirmer et al. 2007), the CFHTLS-Wide (Shan et al. 2012), the HSC survey (Miyazaki et al. 2015), and the KiDS data (Viola et al. 2015).

Summary

In this chapter, weak gravitational lensing has been presented considering both theoretical and observational aspects.

Under the thin-lens approximation and the Born approximation, we see how weak lensing can be presented with two quantities, the convergence κ and the shear γ, from a mass density distribution. We also derive the expected lensing signal for a particular case of cluster lensing.

Then, to reconstruct the lensing signal, an unbiased shear estimator is introduced. To maximize the extraction of cosmological information, we can not be satisfied with two-point statistics, and peak counts seem to be a good candidate to investigate.

Lensing faces numerous observational challenges, such as shape measurement, shape noise, masking, and source redshifts. Also, modelling all physical effects is not trivial. The mass-sheet degeneracy, intrinsic alignment, baryonic physics, and the nonlinear matter spectrum are further issues.

At the end, I provide a review on studies about weak-lensing peak counts, highlighting that methodologies vary as the purposes, cosmology-oriented or cluster-oriented, are different. In the next chapter, I would like to invite readers to address the core issue of the thesis: peak-count modelling.

Chapter V

Peak-count modelling

Overview

Although peak counts are straightforward to compute from a data-processing point of view, predicting them for a given cosmological model and a parameter set is more challenging. The aim of this chapter is to propose a solution to this. First, I will point out the difficulties related to peak-count modelling. Then, the focus will be put on a new method to model lensing peak counts and its pros and cons. The new model will be validated by N-body simulations. Finally, the last section will show a comparison with an analytical model. This chapter corresponds to Paper I and some additional contents.

5.1 Problematic of peak-count modelling

The origin of the problem is that *peak counts are not quantities that can be linked with a simple formula to any known analytical expression.* In the peak-count framework, this analytical support should be the halo mass function that cosmologists derive either from structure collapse theories or from N-body-simulation fits. Then, due to some complexity from halo geometry, the projection effect, shape noise, etc., the link between cosmology and observed lensing peaks is no longer trivial. How to properly model lensing peak counts taking into account realistic survey conditions? This is the question in which cosmologists are interested. Until now, three approaches have been proposed: (1) analytical models, (2) modelling using N-body simulations, and (3) a fast stochastic forward approach developed in this work.

Drawbacks of analytical models

The description of analytical solutions is as follows. Take a noisy convergence map. The idea is to consider galaxy shape noise as a Gaussian random field and lensing signal as a "foreground". In random field theory, the probability to have a local maximum of a given level can be well described as a function of the noise level and the value of the foreground. Therefore, constructing a *peak function* (peak number density as a function of S/N value) becomes possible. This approach has been adapted by Maturi et al. (2010, 2011) and Fan et al. (2010). The differences between two models are subtle. Fan et al. (2010) work directly on a convergence map while Maturi et al. (2010) extract peaks from a shear catalogue using

aperture mass. Fan et al. (2010) define local maxima as peaks while Maturi et al. (2010) count *contiguous areas exceeding a given threshold*. The remaining settings are similar.

In spite of providing explicit expressions for the peak function, analytical models have some harmful drawbacks. When realistic conditions are taken into account, the performance of these models are strongly limited. For example, how to model peak counts when a part of the data is masked? What about the bias from photometric redshifts (photo-z) or the errors from shape measurement? The impact from these effects might not be minor for peak-count observables. A tentative approach for studying masking effects has been done by Liu X. et al. (2014). They improve on the model of Fan et al. (2010) with a two-regime strategy for dealing with masking effects. Nevertheless, modelling realistic survey settings and errors in general still remains unsolved for analytical solutions.

Another drawback is the difficulty to include additional cosmological or astrophysical features. The most intriguing topic among these is intrinsic alignment (IA, see Sect. 4.4). For analytical models, the presence of IA is interpreted as additional noise which is potentially non-Gaussian. This invalidates the peak function obtained from a Gaussian field and thus breaks down the models. An attempt to model the impact from IA on peaks has been proposed by Fan (2007). That work simply considers the Gaussian part of IA, so that the observed variance becomes the sum of the original variance and the corrected one from IA. However, this is strongly unsatisfactory since (1) IA does not necessarily enhance the noise variance and (2) two totally opposite IA variations might lead to the same effect in this consideration. Until now, the impact of IA on WL peak counts has not been studied yet, and no sophisticated solution for IA features has been proposed for analytical peak-count models. Similar to IA, the baryonic effects, influencing the massive core of clusters, could introduce biases to peak-count models. Osato et al. (2015) have examined this effect and have found that the constraint for w_0^{de} can be offset up to 0.061, which is $\sim 1\sigma$ for a HSC-like 1400-deg^2 survey. Neglecting these extensions could result in large systematic biases. Taking them properly into account is crucial for peak studies.

The last disadvantage is the fact that analytical models count on external simulations to estimate statistical uncertainty. The propagation of errors from observational effects, such as photo-z and masking, and scatters due to some stochastic processes, for example halo triaxiality and concentration dispersion, cannot be quantified by analytical calculations alone in general. In cosmological contexts, models of this kind require N-body simulations for estimating the covariance matrix of observables under the assumption of a Gaussian likelihood. At the end of the day, the total process from the theoretical basis to parameter constraints still requires considerable computational resources. As a result, analytical models seem to be neither accurate nor efficient.

What about our old good friend, N-body simulations?

Alternatively, a straightforward approach for modelling peak counts is to use N-body simulations. These runs simulate structure formation of the Universe throughout the time evolution, and provide simultaneously the prediction and the statistical variability of cosmological observables. Survey conditions and additional astrophysical features can be taken into account directly, since this is a forward modelling approach. However, N-body simulations have an embarrassing feature which is their large computation time. Depending on computational resources and simulation settings, the time cost for creating a N-body lightcone can vary between a day and several months. Repeating this process for a wide range of cosmological

parameters is extremely expensive, especially when the number of parameters increases. This so-called *curse of dimensionality* makes the grid-point evaluation of likelihood using N-body simulations practically impossible.

Nowadays, different sampling techniques, such as Markov Chain Monte Carlo (MCMC) and population Monte Carlo (PMC), are widely used for accelerating the constraining process. With these techniques, the reasonable number of parameter sets to evaluate is no longer an exponential function of the parameter dimension. However, the time cost for N-body runs still remains very harmful. Cosmologists usually need to make a tradeoff between several factors: the dimension of parameters, the prior range, the resolution of the parameter space, statistical precisions, the resolution of simulations, and computing power. For example, the study of Liu J. et al. (2015a) using N-body simulations adopts a strategy with a large 3D $(\Omega_{\mathrm{m}}, \sigma_8, w_0^{\mathrm{de}})$ prior but a poor resolution and a poor statistical precision (one run per parameter set).

The drawbacks described above motivates the development of a new model in this thesis work. I will present this new model in the next section.

5.2 A new model to predict weak-lensing peak counts

5.2.1 Description

The idea of the new model is to replace the time-costly N-body physical process by a "short-cut": an accelerated process which provides the same observables. A possible solution is semi-analytical computations under some assumptions. In this case, the new model preserves some characteristics of N-body runs, forward and probabilistic, and gets rid of the considerable time cost. This approach is called *fast stochastic forward modelling* (FSF modelling). Clearly, this principle is not reserved only for WL peak counts. As long as the shortcut can be constructed, applying this fast modelling in other contexts is straightforward.

The fast stochastic forward model to predict WL peak counts developed in this thesis adopts steps as follows:

- sample halos from a mass function;

- assign density profiles, randomize their position;

- compute the projected mass, add noise;

- make maps, create peak catalogues.

The two first steps are called "fast simulations" in Paper I, because these processes correspond to the N-body physical process. Actually, a more intuitive name for fast simulations is the "bubble model" . Since in most cases halos are considered to be spherical, a fast simulation box is actually a bubble chamber, or more precisely, a bubble lightcone with respect to a redshift distribution. The requirement for creating these is a mass function and a halo mass profile. Readers may notice that sampled halos might overlap in the 3D simulated boxes. We do not look for detecting or excluding these cases. The projected mass is computed by summing up contributions from each halo on the line of sight.

The model is implemented by our public code CAMELUS [♯1], standing for *Counts of Amplified Mass Elevations from Lensing with Ultrafast Simulation.* The cosmological computation

---

[♯1]Available at http://github.com/Linc-tw/camelus/.
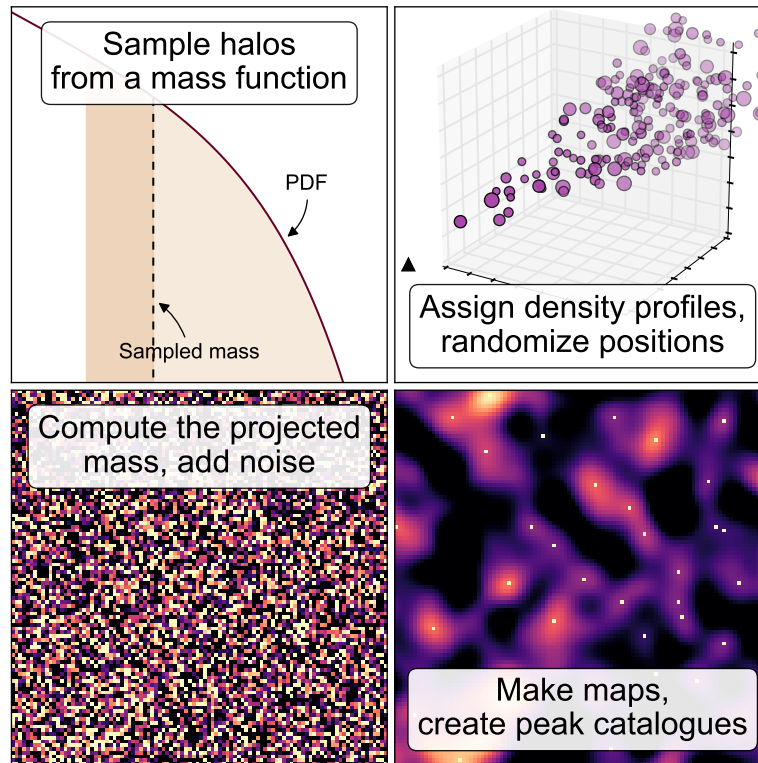
FIGURE 5.1: Illustration of various steps of our model to predict weak-lensing peak counts.

FIGURE 5.2: Logo of CAMELUS.

in CAMELUS is performed by NICAEA (Kilbinger et al. 2009) [♯2].

By construction of the fast simulations, one may find out that this model is based on two major hypotheses:

- diffuse, unbound matter, for example cosmological filaments, does not significantly contribute to peak counts;

- the spatial correlation of halos has a minor influence and can be ignored for peak counts.

Implicitly, this framework also admits the universality of the mass function and halo models. The first assumption is motivated by the fact that unbound mass is less concentrated than bound structures. The second one is supported by Marian et al. (2010), who have proven that correlated structures affect number and height of peaks by only a few percent. Similarly, Kainulainen & Marra (2011) have found that assuming a stochastic distribution of halos can lead to accurate predictions of the convergence probability distribution function (PDF). These studies motivate the use of halo one-point-only statistic in our model, while a detailed examination of above will be presented in Sect. 5.3.

[♯2] Available at http://www.cosmostat.org/software/nicaea/.

Readers may understand now why fast simulations are good candidates to substitute the N-body physical process. Actually, most peaks with S/N values (noted as ν) lower than 3–4 are dominated by galaxy shape noise (see Fig. 5.4), while high peaks are basically true structures: either a single massive cluster or several halos on the same line of sight, and Yang et al. (2011) suggest that the first scenario is more plausible. They have found that $\sim 74\%$ of peaks with $\nu \geq 4.8$ are in the single-halo case. As a result, cutting off unbound objects and low-mass halos will not modify high-peak counts, and this helps drastically reducing the computation time.

On the other hand, although low peaks are dominated by noise, Yang et al. (2013) have shown that they constrain cosmology better than high peaks due to the fact that cosmological variations are larger than statistical fluctuations. However, since our model introduces a lower mass cutoff during the sampling step, the missing low-mass halos might create a substantial impact on the regime of low peaks. Due to this reason, low-ν peaks are not included in the first consideration.

5.2.2 Advantages

What are the interests of an FSF model like ours? What do we gain by taking the fast simulation? The improvements from it can be summarized by three characteristics: fast, flexible, and full PDF information.

Fast On a single-central-processing-unit (single-CPU) machine, it only requires several seconds for creating a 25-deg^2 field. This is without using parallel computing such as message passing interface (MPI) or graphics processing unit (GPU) programming. The main cost is the computation of the projected mass, or nonlinear filtering method if it is used.

Flexible Including observational effects, such as masking and photo-z errors, and additional features, such as IA and baryonic feedbacks, is straightforward, since the model adopts a forward approach. A detailed discussion for possible extensions is presented in Sect. 5.4.

Full PDF information A stochastic model provides naturally an empirical PDF of observables, which is supposed to contain all statistical information. This requires a large amount of model realizations, difficult to carry out for N-body runs. The fact that a FSF model is fast makes the PDF reconstruction feasible in practice.

With these three characteristics, constraining parameters becomes less restricted. For example, when the parameter dimension remains small, a grid-point evaluation of the likelihood in parameter space with high resolution will be feasible. Another example is the hypothesis of the constant covariance that almost all studies make for their Gaussian likelihood. A FSF model can easily test the impact of the cosmology-dependent-covariance effect (CDC, see Sect. 6.2). It is also possible to apply non-parametric constraint methods such as p-value tests or approximate Bayesian computation (ABC). A more detailed study will be presented in Chaps. 6 and 7.

5.3 Validation by N-body simulations

5.3.1 Description of N-body and ray-tracing simulations

Putting calibration of systematics aside, here we attempt to validate our FSF modelling by comparing it to N-body simulations. The N-body simulations used are the simulations from Dark Energy Survey (DES) Blind Cosmology Challenge, called "Aardvark". They have been generated by LGADGET-2, a DM-only version of GADGET-2 (Springel 2005). The cosmological parameters for Aardvark form a WMAP-like ΛCDM cosmology, with $\Omega_{\mathrm{m}} = 0.23$, $\Omega_{\Lambda} = 0.77$, $\Omega_{\mathrm{b}} = 0.047$, $\sigma_8 = 0.83$, $h = 0.73$, $n_{\mathrm{s}} = 1.0$, and $w_0^{\mathrm{de}} = -1.0$. The DM halos in Aardvark were identified using the ROCKSTAR friends-of-friends (FOF) finder code (Behroozi et al. 2013).

Ray-tracing simulations for Aardvark have been performed with CALCLENS (Becker 2013). This is an algorithm yielding multi-plane lensing computation on a curved sky. Galaxies of Aardvark have been generated with ADDGALS (by M. Busha and R. Wechsler).

I use a HEALPix (Górski et al. 2005) patch with `nside` $= 2$ of the halo catalogue, which is $860 \ \mathrm{deg}^2$ (large field), and a subpatch with `nside` $= 8$ of the galaxy catalogue and ray-tracing data, which is only $54 \ \mathrm{deg}^2$ (small field). In order to use as large field of view as possible, the map has been created in HEALPix pixels. The `nside` of each pixel is 16384. Fast simulations from our model have also been generated in the same fields (see Sect. 5.3.2).

5.3.2 Methodology

Comparison design

In order to understand the impact from different hypotheses mentioned in Sect. 5.2.1, two intermediate steps have been included for comparison. This ends up with four different settings in total:

- Case 1: full N-body runs,

- Case 2: replacing N-body halos with NFW profiles of the same mass,

- Case 3: randomizing angular positions of halos from Case 2,

- Case 4: our model.

Readers can notice that, first, halos of Case 1 are from the N-body process and recovered by a FOF or spherical overdensity (SO) finder, while halos of Cases 2, 3, and 4 follow NFW profiles. Second, the position correlation between halos is conserved in Cases 1 and 2, while no correlation is retained for Cases 3 and 4. Last, Cases 1, 2, and 3 follow the same mass function as N-body simulations, while the one of Case 4 is an analytical model. At the end of the day, by comparing Cases 1 and 2, one studies the impact from unbound matter and halo asphericity. The comparison between Cases 2 and 3 should be interpreted as the impact of halo position correlations. And the difference between Cases 3 and 4, if there is any, corresponds to verify if the mass function from the N-body is well described by the analytical model.

Settings for Case 1

This is tested in a simplified scenario, without masking, IA, and baryonic physics. Also, only a slice of sources is considered. Such a configuration has widely been adopted in the literature for preliminary studies, including Hamana et al. (2004), Fan et al. (2010), and Yang et al. (2011).

For Case 1, I take galaxies with redshifts between 0.9 and 1.1. Lensing in this case has already been described in Sect. 5.3.1. No inversion is needed since κ has been directly obtained from ray-tracing. To make a grid map, I use the convergence values sampled by galaxies at this source slice, and carry out a bilinear interpolation. The "truth" map is then obtained by evaluating interpolated field values on HEALPix pixel positions. In reality, this "truth" can never been recovered because of galaxy shape noise, thus the comparison should always be done between noisy fields.

Shape noise can be modelled as a Gaussian random field $n(\boldsymbol{\theta})$ added to convergence, giving the noisy field $\kappa_n(\boldsymbol{\theta})$ as

$$\kappa_n(\boldsymbol{\theta}) = \kappa(\boldsymbol{\theta}) + n(\boldsymbol{\theta}). \tag{5.1}$$

For $n(\boldsymbol{\theta})$, the noise level depends on the smoothing function. If we bin galaxies into map pixels, the smoothing function is just a top-hat filter based on the pixel support. In this case, the noise level σ_{pix} for a pixel of size A_{pix} is given by (Van Waerbeke 2000)

$$\sigma_{\mathrm{pix}}^2 = \frac{\sigma_\epsilon^2}{2} \frac{1}{n_{\mathrm{gal}} A_{\mathrm{pix}}}, \tag{5.2}$$

where σ_ϵ is intrinsic ellipticity dispersion and n_{gal} is galaxy number density. The intrinsic dispersion is set to $\sigma_\epsilon = 0.4$ which corresponds to a CFHTLenS-like survey (Fu et al. 2008), and the value of n_{gal} will be discussed further in this section.

Of course, galaxy binning is not necessary in general. The advantage of binning is that when the grid is Cartesian, the computation can be significantly sped up using the fast Fourier transform (FFT). The speed is one of the priority of our algorithm, so this setting is always used. This is why the same spirit is kept even if the grid is not Cartesian here.

After adding random noise to pixels, the map is smoothed with a Gaussian kernel $W(\boldsymbol{\theta})$ defined as

$$W(\boldsymbol{\theta}) = \frac{1}{\pi \theta_{\mathrm{G}}^2} \exp\left(-\frac{\theta^2}{\theta_{\mathrm{G}}^2}\right), \tag{5.3}$$

with $\theta_{\mathrm{G}} = 1$ arcmin. One can then defined the smoothed convergence as

$$K_N(\boldsymbol{\theta}) \equiv (\kappa_n * W)(\boldsymbol{\theta}) = \int \mathrm{d}^2\boldsymbol{\theta}' \ \kappa_n(\boldsymbol{\theta} - \boldsymbol{\theta}')W(\boldsymbol{\theta}'), \tag{5.4}$$

where κ_n and K_N denote noisy fields before and after smoothing respectively. In the validation study, this is performed by computing direct convolution. After smoothing, the noise level is expected to be (Van Waerbeke 2000; Sect. 8.2.1)

$$\sigma_{\mathrm{noise}}^2 = \frac{\sigma_\epsilon^2}{2} \frac{1}{2\pi n_{\mathrm{gal}} \theta_{\mathrm{G}}^2}. \tag{5.5}$$

The S/N of each pixel is then defined as $\nu \equiv K_N/\sigma_{\mathrm{noise}}$. This quantity has been used for selecting peaks.

In practice, after obtaining the interpolated "truth" map, I add a random sampled value from $\mathcal{N}(0, \sigma_{\mathrm{pix}}^2)$ to each pixel as noise. Smoothing is done by computing a direct convolution, and peaks are identified as pixels such that the S/N is larger than all eight neighbors. Some most outer pixels are abrogated to avoid the border effect. The width of this area corresponds approximately to $2.2 \times \theta_{\mathrm{G}}$, which is three times the standard deviation of W. The field size for Case 1 is 54 deg^2.

Settings for Case 4, a "minimal pipeline"

I would like to skip Cases 2 and 3 for the moment and come back to them later. For Case 4, peak counting is processed with a "minimal pipeline". I first generate fast simulation boxes. These boxes are actually composed of redshifts slices. In this validation test, slices are set to 10 equal bins from $z = 0$ to 1. Halo masses are sampled from the mass function proposed by Jenkins et al. (2001, Eq. (3.81)). The sampling mass range is $[M_{\mathrm{min}}, M_{\mathrm{max}}] = [10^{12}, 10^{17}]~M_\odot/h$, and this implies that the total mass M_{tot} of each slice is not $\rho_{\mathrm{crit}} \cdot \Omega_{\mathrm{m}} \cdot V$, but

$$M_{\mathrm{tot}} = V \int_{M_{\mathrm{min}}}^{M_{\mathrm{max}}} \mathrm{d} \log M \cdot M \cdot \frac{\mathrm{d} n(z, <M)}{\mathrm{d} \log M}, \qquad (5.6)$$

where V is the volume of the slice. Halo profiles are NFW ones (inner slope $\alpha = 1$) truncated at r_{vir}. They are parametrized by mass and concentration. Here, I assume the mass-concentration relation as Eq. (3.93) and set $c_0 = 8$ and $\beta_{\mathrm{NFW}} = 0.13$.

Sources are considered to be regularly spaced on a HEALPix grid and fixed at a single redshift $z_{\mathrm{s}} = 1$. The HEALPix grid corresponds either to the halo catalogue patch (large field) or to the galaxy catalogue patch (small field). In both cases, galaxies are located at the center of pixels with $\texttt{nside} = 16384$. This results in $A_{\mathrm{pix}} = 0.046~\mathrm{arcmin}^2$ and $n_{\mathrm{gal}} = 21.7~\mathrm{arcmin}^{-2}$. This is also the value of n_{gal} for Case 1.

After setting lenses and sources, the lensing signal is calculated. For Case 4, this is done by summing up the projected mass along the line of sight following

$$\kappa_{\mathrm{proj}}(\boldsymbol{\theta}, w_{\mathrm{s}}) \equiv \sum_{\mathrm{halos}} \kappa_{\mathrm{halo}}(\boldsymbol{\theta}, w_\ell, w_{\mathrm{s}}) \qquad (5.7)$$

where κ_{halo} is given by Eqs. (4.62) and (4.63), and then I subtract the mean over the field at the end. The reason for this is that the projected mass from halos is always positive, while the true convergence can be negative in underdense regions. Recall that in Sect. 4.2.1, we have already seen that Eq. (5.7) can be considered as the convergence only under an important condition: the "empty space" actually has the mass density of the background. Consequently, the calculation using Eq. (5.7) without any correction leads to a κ map from an universe in which the total mass is larger than what it should be for the input Ω_{m} value. In our model, instead of subtracting this additional mass, I simply subtract the mean of κ_{proj}, i.e. considering $\kappa(\boldsymbol{\theta}) = \kappa_{\mathrm{proj}}(\boldsymbol{\theta}) - \overline{\kappa}_{\mathrm{proj}}$. The philosophy behind is to make a zero-mean convergence field an approximation of the true convergence field, which arises from a zero-mean density contrast on average. After obtaining the noiseless map, I add noise, smooth, and count peaks on S/N as in Case 1.
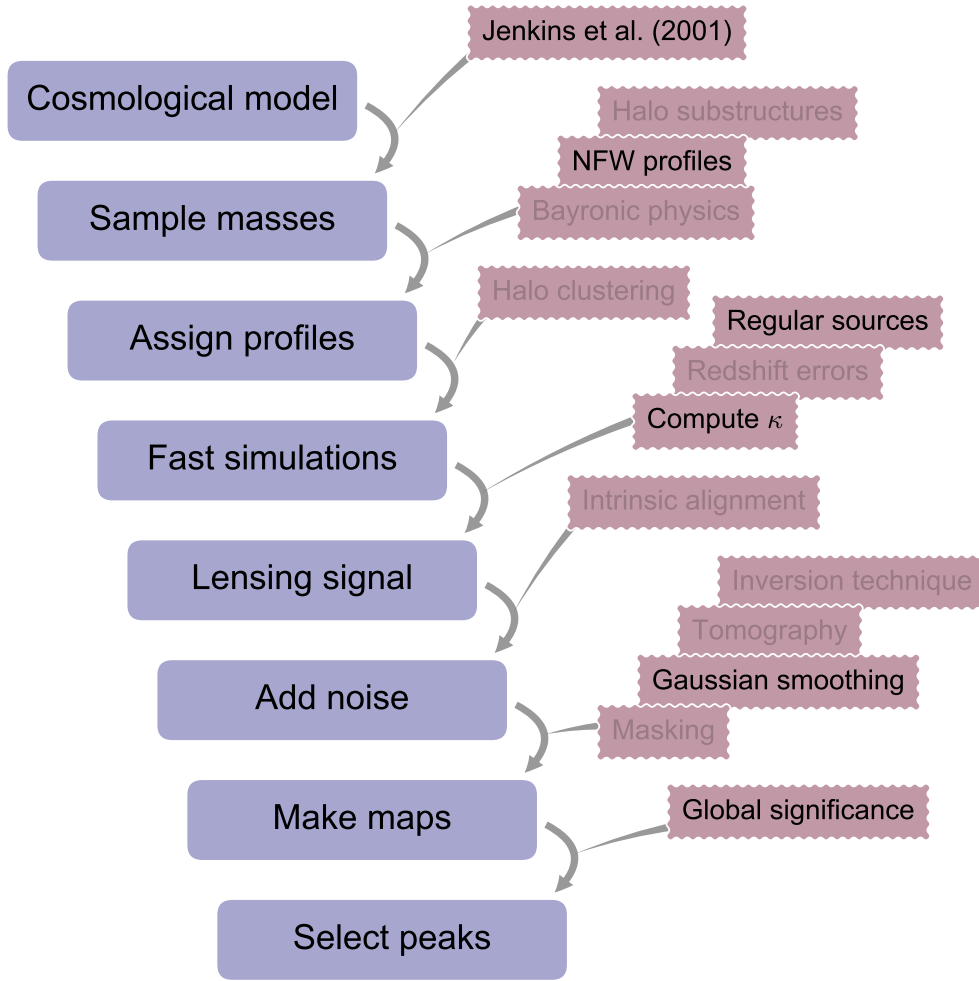
FIGURE 5.3: Diagram illustrating the processing pipeline for the validation test. Effects which are not considered are made faint. A more detailed description for all variations can be found in Sect. 5.4.

Settings for Cases 2 and 3

For Cases 2 and 3, the processing is similar to Case 4 apart from fast simulations. In these two cases, I do not need to generate halos, but profiles are analytical, so the new lensing signal is computed by following Case 4. For Cases 2, 3, and 4, peaks are selected both from the large field and the small field, since these cases are not limited by the existing galaxy catalogue with ray-tracing information.

It is worth clarifying that $n_{\mathrm{gal}} = 21.7$ arcmin^{-2} is used for all cases. Thus, $n_{\mathrm{gal}} A_{\mathrm{pix}}$ is 1 automatically. Even if galaxy density of the sources in the redshift slice of [0.9, 1.1] from Case 1 is lower than 21.7 arcmin^{-2}, since the aim is to test different lens conditions under the same source configuration, the same noise level should be kept everywhere. I also consider that galaxies are distributed uniformly, so that n_{gal} is independent from position. As a result, here, ν is just a scaling of K_N. However, in general conditions, especially when masking is taken into account, n_{gal} becomes a local quantity which varies from one point to another. In this case, evaluating noise locally and counting peaks on true S/N would be more authentic than scaling all K_N to ν with the same factor.

Parameter	Symbol	Value
Lower sampling limit	$M_{\min}$	$10^{12} \ M_{\odot}/h$
Upper sampling limit	$M_{\max}$	$10^{17} \ M_{\odot}/h$
Number of halo redshift bins	-	10
NFW inner slope	α	1
M-c relation amplitude	c_0	8
M-c relation power law index	β_{NFW}	0.13
Source redshift	z_{s}	1
Intrinsic ellipticity dispersion	σ_ϵ	0.4
Galaxy number density	n_{gal}	21.7 arcmin^{-2}
Pixel size	θ_{pix}	0.215 arcmin
Gaussian kernel size	θ_{G}	1 arcmin
Noise level in a pixel	σ_{pix}	0.283
Noise level after smoothing	σ_{noise}	0.0242
Small field size	-	53.7 deg^2
Large field size	-	859 deg^2

TABLE 5.1: List of parameter values adopted in this validation study.

For the small field, 8 independent noise maps are generated for analysis, and all cases with random processes (random position for Case 3 and fast simulation for Case 4) are carried out with 8 realizations. This ends up with averaging the result over either 8 or 64 peak histograms. For the large field, I generate 4 noise maps and 4 realizations instead, so the average is over 4 or 16 histograms. However, one should also notice that the large field is 16 times larger than the small field.

5.3.3 Results

Small field, low-ν regime

Figure 5.4 shows the comparison between four cases on the small field. Only halos located in the 54-deg^2 subpatch are taken into account for analysis. The x-axis represents peak S/N, denoted as ν. The y-axis stands for the peak function $n_{\mathrm{peak}}(\nu)$, which are histogram counts divided by the binwidth and the field area. We can discover that in the regime of low peaks with $\nu \leq 3.75$, $n_{\mathrm{peak}}(\nu)$ does not vary much between four cases. This is not surprising because noise is dominating in this range, as shown by the noise-only peak-count histogram (purple dashed lines in Fig. 5.4).

The second observation from the low-ν regime is non-additivity. Combining peaks from cosmological structures and noise does not give the total count from the noisy field. This is because a pixel needs to "cooperate" with neighbors to become a local maximum and summing up two fields yields an ambiguous effect for this "cooperation". Peaks with $\nu \leq 2.75$ in Fig. 5.4 show that appending structures to the noise field even decreases the peak counts.

The lower panel of Fig. 5.4 gives the deviation of each case compared to Case 1. It shows that there exists a systematic overcount of 10% in the low-ν regime. Several explanations for its origin are possible. First, it could come from the κ-mean subtraction. For example, if
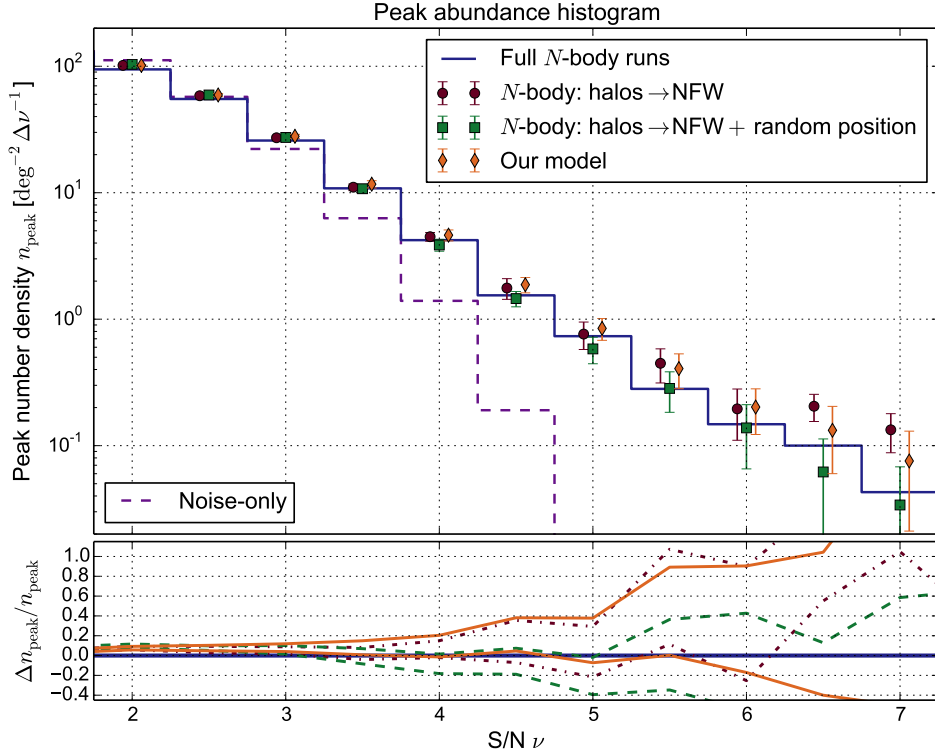
FIGURE 5.4: Comparison of the peak abundance from different cases evoked in Sect. 5.3.2. On the upper panel, blue solid lines: full N-body runs (Case 1); dark red circles: replacement of halos by NFW profiles (Case 2); green squares: replacement of halos by NFW profiles and randomization of halo angular positions (Case 3); orange diamonds: fast simulations, corresponding to our model (Case 4); purple dashed line: peaks from noise-only maps. On the lower panel, I display the upper and lower limits of error bars shifted with regard to Case 1. This refers to the standard deviation over 4 maps (dark red dash-dotted lines for Case 2) or 16 maps (green dashed lines for Case 3, orange solid lines for Case 4). The field of view is 54 deg^2.

NFW profiles truncated at $2r_{\mathrm{vir}}$ instead of r_{vir} are used, then $\overline{\kappa}_{\mathrm{proj}}$ would enhance and peaks will have a lower S/N. Second, it could be the fact that a lower limit M_{min} has been defined for halo mass sampling. Small halos might have an impact on low-peak counts. Finally, it could also be an effect caused by profiles, as I will discuss in the comparison Case 1-Case 2.

Small field, high-ν regime, comparison Case 1-Case 2

As the S/N increases, the differences between cases become significant. We can observe that in the high-ν regime, replacing profiles enhances the peak counts while randomizing position introduces an opposite effect of a similar order of magnitude. Since peak counts at $\nu \gtrsim 4$ can barely contributed by unbound matter or small halos, the enhancement from Case 2 might be mainly caused by halo asphericity or triaxiality.

Another explanation comes from the mass-concentration relation (M-c relation). Simulations have shown that for NFW profiles, M and c are not linked with a tight relation. On the contrary, the scatter of $c(M)$ is rather large (e.g. Fig. 5.5). At a fixed M, the larger c is, the higher the peak is observed. Thus, the modelling of the M-c relation could have an impact on the comparison between Cases 1 and 2. Figure 5.6 illustrates a simple test of dependency of our model on c_0. Note that the runs are done on the large field, whereas the blue lines still stand for Case 1 on the small field. While all the remaining settings are identical to
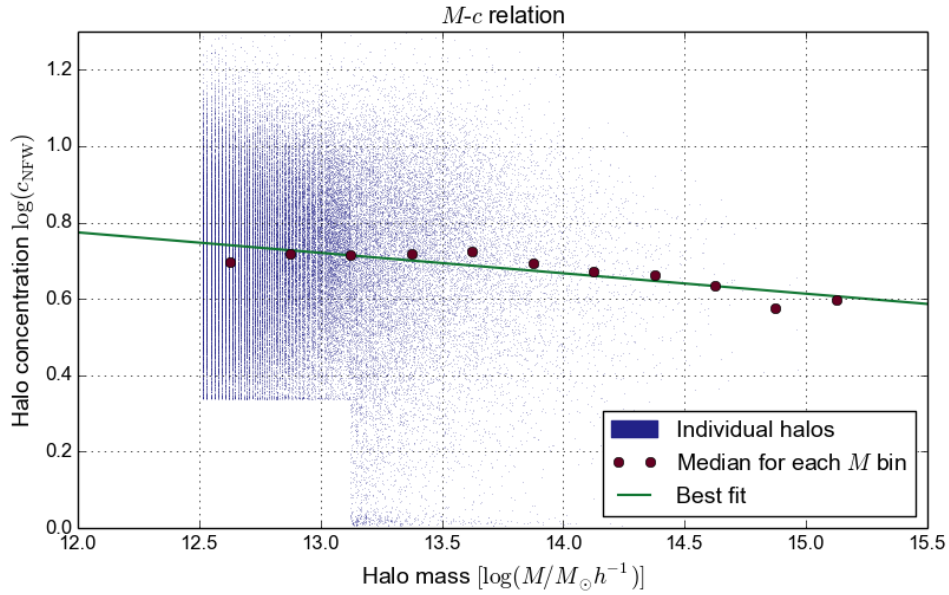
FIGURE 5.5: Scatter of the concentration for the Aardvark halos at redshift between 0.4 and 0.5. The concentration c is derived from the ratio $r_{\mathrm{vir}}/r_{\mathrm{s}}$, where r_{s} is given by fitting an analytical NFW profile to N-body particles in halos. For each mass bin in the logarithmic space, the value of the median of the concentration is drawn with dark red circles, and the best fit of the M-c relation using Eq. (3.93) is presented by the green line.

Case 4, one can clearly observe different peak abundance from different c_0 values. Thus, it is indispensable to include halo parameters such as c_0 and β_{NFW} into the parameter set for future works focusing on constraints. In the validation test, the value of $c_0 = 8$ is suggested by fitting with Aardvark runs, assuming Eq. (3.93) as the M-c relation. However, this fit suffers from some technical limits: a cutoff at $\log(c) \approx 0.32$ for $M \lesssim 10^{13.2}\ M_\odot/h$ and some excess at $c = 1$ for $M \gtrsim 10^{13.2}\ M_\odot/h$, as we can observe from Fig. 5.5.

One possible way to improve the bias induced by the M-c relation is to take the individual concentration of N-body halos into account. Recall that in Case 2, only masses are retained from N-body runs. However, with this strategy, the model loses its universality. For future works, it could be interesting to set up a new intermediate case as current Case 2 with N-body concentrations. In this case, the impact from the M-c relation can be tested.

Small field, high-ν regime, comparison Case 2-Case 3

By comparing Cases 2 and 3 from the upper panel of Fig. 5.4, I discover that position randomization decreases peak counts by 10% to 50% in the high-ν regime ($\nu \geq 3.75$). The explanation of this is that decorrelating angular positions breaks down the two-halo term, so that halos overlap less in projection on the field of view and decreases high-peak counts. Recall that Yang et al. (2011) have shown that high peaks ($\nu \geq 4.8$) are mainly contributed by one single halo. The same study has also shown that 12% of total high-peak counts are derived by multiple halos. Apparently, these 12% are not negligible, and this number agrees with the difference between Cases 2 and 3 within the bin of $4.75 \leq \nu \leq 5.25$, the dominant bin for peaks with $\nu \geq 4.8$.
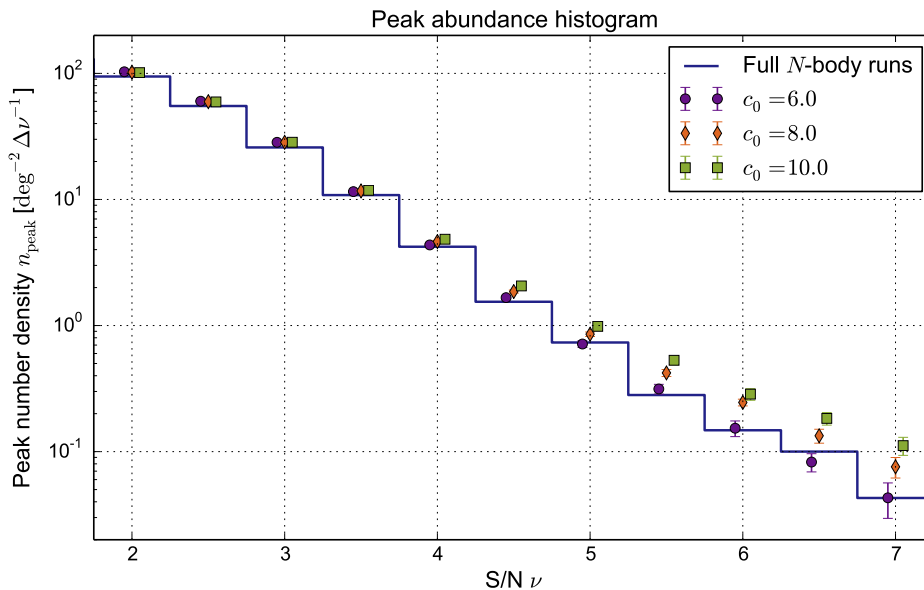
FIGURE 5.6: Peak abundance from our model for different c_0 values. This test has been carried out on the large field, so that orange diamonds are the same runs as Case 4 from Fig. 5.9, which will be mentioned later. This figure suggests that halo parameters and cosmological parameters can be highly degenerated together. The results from N-body runs (Case 1, blue solid lines), are only indicative, since they are derived from the small field and the fluctuation is large in principle.

Small field, high-ν regime, comparison Case 3-Case 4

The last comparison from Fig. 5.4, between Cases 3 and 4, studies the impact of the mass function modelling. The result shows that more peaks have been found using the analytical model of Jenkins et al. (2001), and the excess somehow compensates for the deficit from randomization. To explain this excess, we can take a look at Fig. 5.7 which draws both mass functions at four different redshifts. While both mass functions agree well with each other, some discrepancies are still worth being pointed out.

First, at low redshift, N-body runs have more halos than the analytical model by a factor of a few. However, low redshift means poor lensing efficiency for halos, so they should not have strong impact. In addition, the number of peaks varies in the opposite way to the number of halos from Case 3 to Case 4. This is then not the origin of the excess.

The second difference is that at low M, the N-body mass function has an irregular tail for the reason of limited resolution, and depending on redshift, this irregular tail either contains more or less halos than fast simulation boxes. This hypothesis does not provide satisfactory explanation either because low-mass halos have a minor influence on high peak counts. To justify this, Fig. 5.8 shows the result from our model with three different values of $M_{\min}$. Some hints suggest that deficits may occur for peaks with $3.25 \leq \nu \leq 4.75$ when $M_{\min} = 10^{13} \, M_\odot/h$, but for $M_{\min} \lesssim 10^{12} \, M_\odot/h$, high-peak counts are not so sensitive to $M_{\min}$. As a result, the irregular tail of the N-body mass function can be ignored.

The last difference is that at high redshift ($z = 0.65$ and 0.95), the analytical model predicts more halos. Although the quantitative results between the halo excess ($\lesssim 50\%$ for high M and high z) and the peak excess (50%–100% for $\nu \geq 4.25$) do not match perfectly, this still remains the most probable explanation for the Case 3-Case 4 difference. Since the same massive halo can result in different S/N due to the statistical fluctuation, there is no one-to-
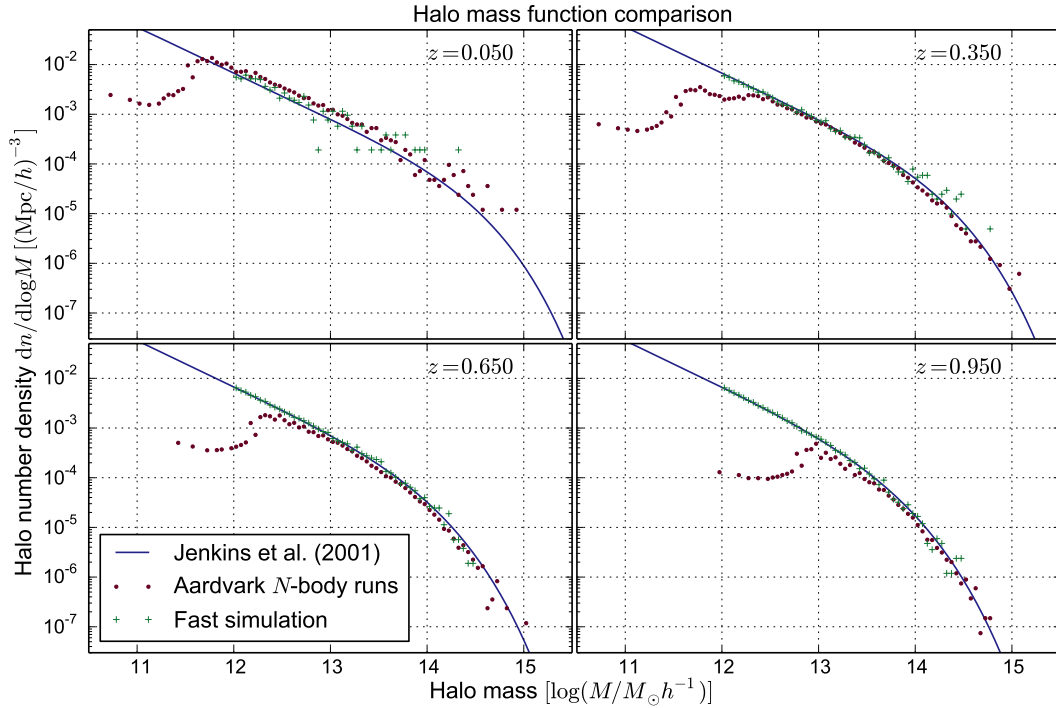
FIGURE 5.7: Mass function comparison between an analytical model and simulations at different redshifts. At low redshift, N-body runs count more halos than the analytical model. At high redshift, this effect is slightly reversed. In all cases, the low-mass part disagrees due to the mass resolution in the simulations.

one relation between halo masses and peak heights. Therefore, predicting the quantitative impact of mass function on peak abundance is difficult. As a result, the difference between Cases 3 and 4 would be mainly the consequence of the mismatch between two mass functions.

Large field

The end of this section focuses on the same comparison on the large field. Figure 5.9 shows Cases 2, 3, and 4 performed on the large field together with the Case 1 from the small field. The difference between sizes of fields is a factor of 16, so Case 1 (blue lines) is only indicative in Fig. 5.9 since the uncertainty of blue lines is comparable to error bars in Fig. 5.4. Nevertheless, the result, with a weaker statistical uncertainty, confirms the all observations taken from Fig. 5.4. It suggests that Cases 1 and 3 are in a very good agreement.

Overall, the model works very well for $\nu \leq 4$. On the small field, systematic biases are smaller than the statistical uncertainty for all ν. Although the validation study has shown that not all biases are not entirely understood, the variation from different cosmological predictions are found to be larger than biases (see Sect. 5.5), hence are able to perform model selection. Also, in the next section, reader are going to see how our model can be improved. As far as the absolute bias is concerned, a larger set of N-body simulations would be required in the future.
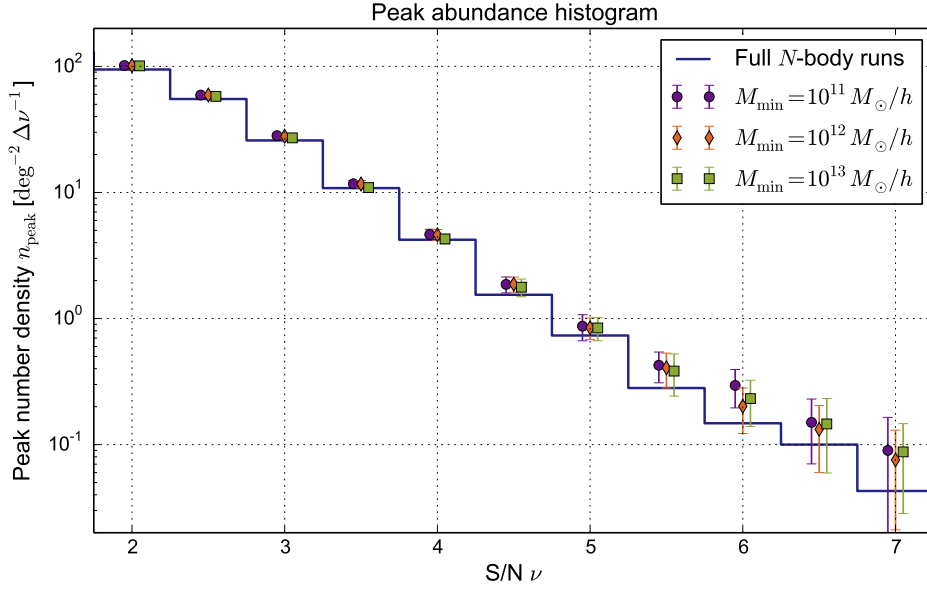
FIGURE 5.8: Peak abundance from our model for different $M_{\min}$ values. This test has been carried out on the small field, because the lower $M_{\min}$ is, the more halos present in fast simulations, which make the computation intractable. Orange diamonds are the same runs as Case 4 from Fig. 5.4, while blues lines stand for Case 1. Without quantitative comparisons, peak counts seem to be insensitive to low-mass halos.

5.4 Model extensions

Thanks to the flexibility of our model, a wide range of options or extensions for different steps is available. These extensions could help us improving the physical modelling and extracting cosmological information in a more optimal way. Below is a non-exhaustive list of extensions that one may include in our model (see also Fig. 5.10).

Mass function   The mass function can be replaced by any distribution of our interests. Conversely, our model can also be used to discriminate mass function models. Nevertheless, this is not the aim of this thesis. In this thesis, the default input mass function is the one from Jenkins et al. (2001).

Halo structures & baryonic feedback   Throughout this work, the NFW profiles (see Sect. 3.5.1) have been chosen for halos. However, other density profiles such that the projected density is known are applicable. Therefore, providing an analytical expression of the projected mass would be the key for using the Einasto profile (Einasto 1965) or elliptical profiles. The later one refers to studies related to halo triaxiality where the privileged direction of halos are taken into account. Beyond the dark-matter-only (DM-only) profiles, those offered by baryonic feedback (Yang et al. 2013; Osato et al. 2015) and substructures inside the halos could also be included in studies.

Halo clustering   It is possible not to break down the halo spatial correlation. The halo clustering information can be generated by using some fast algorithms such as PTHALOS (Scoccimarro & Sheth 2002), PINOCCHIO (Monaco et al. 2002, see also Heisenberg et al. 2011), and remapping LPT (Leclercq et al. 2013). However, this is not the aim of the thesis. Zero angular correlation for halos is assumed throughout this work.
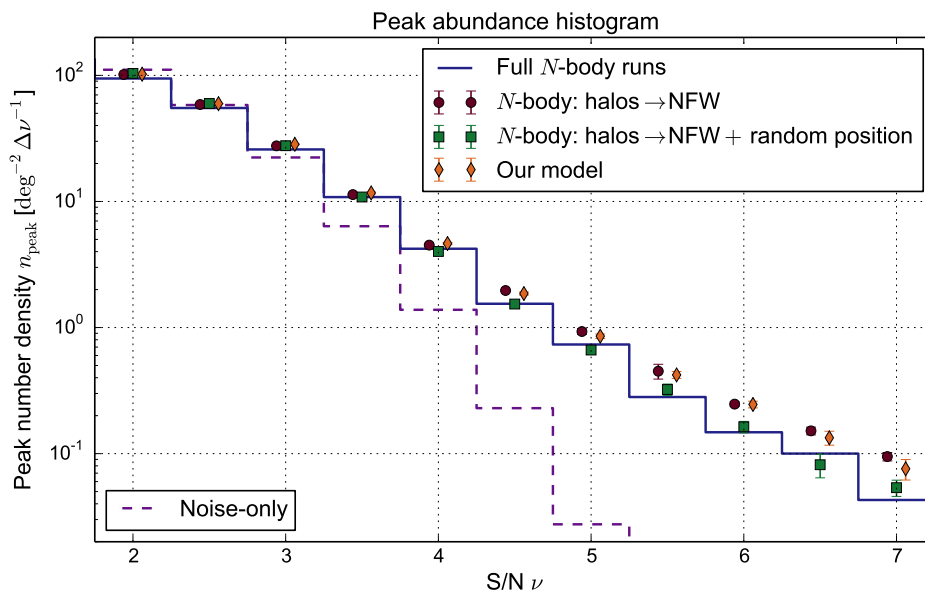
FIGURE 5.9: Similar to Fig. 5.4, comparison between four cases on the large field. Here, Case 1 is only indicative since it is still the result from the small field.

Galaxy redshift, redshift errors, tomography   One can either suppose a constant source redshift or generate sources from a distribution law. If a model for photo-z or other redshift errors is provided, our forward model generates without any difficulties a series of realistic observed redshifts for galaxies. To better extract cosmological information, performing tomographic studies is also feasible.

Galaxy shape noise & intrinsic alignment   Another extension for our model is the noise model. When isotropy for galaxy shape orientation is assumed, Gaussian noise is added. Beyond Gaussian noise, some physically-motivated models for IA can be added easily. In this case, one needs to identify pairs of galaxy and hosted halo. The galaxy orientation can be provided by an IA model depending, on its type and the distance to the halo center.

Inversion & filtering   A large number of options are available for our model to make a mass map. For example, using a κ-peak approach, one could compute directly the convergence signal without worrying about the inversion problem (see Sect. 4.4), while a more realistic way would be simulating shears and using inversion techniques (Kaiser & Squires 1993, Seitz & Schneider 1995, etc.) to get convergence maps. Besides, one can also adopt an aperture-mass approach: convolving the shear field with a zero-mean filter. By definition, the aperture mass is already a smoothing, while both κ-peak methods require in addition filtering techniques, linear or nonlinear ones, to reduce noise (see further Chap. 8). For all three modelling approaches, taking masking into account is not necessarily trivial. Reducing the effective survey area, determining near-mask areas with a higher noise level (Liu X. et al. 2014), or filling missing data with inpainting (method: Pires et al. 2009b, data application: Jullo et al. 2014), are different options.

S/N determination   In most studies, the noise level in S/N is a global value derived from the whole survey, which yields a global significance. However, in reality, depending on mask and the galaxy spatial distribution, the local noise level is not uniform. Taking the local significance into account would make the modelling more accurate (see also Chap. 8).
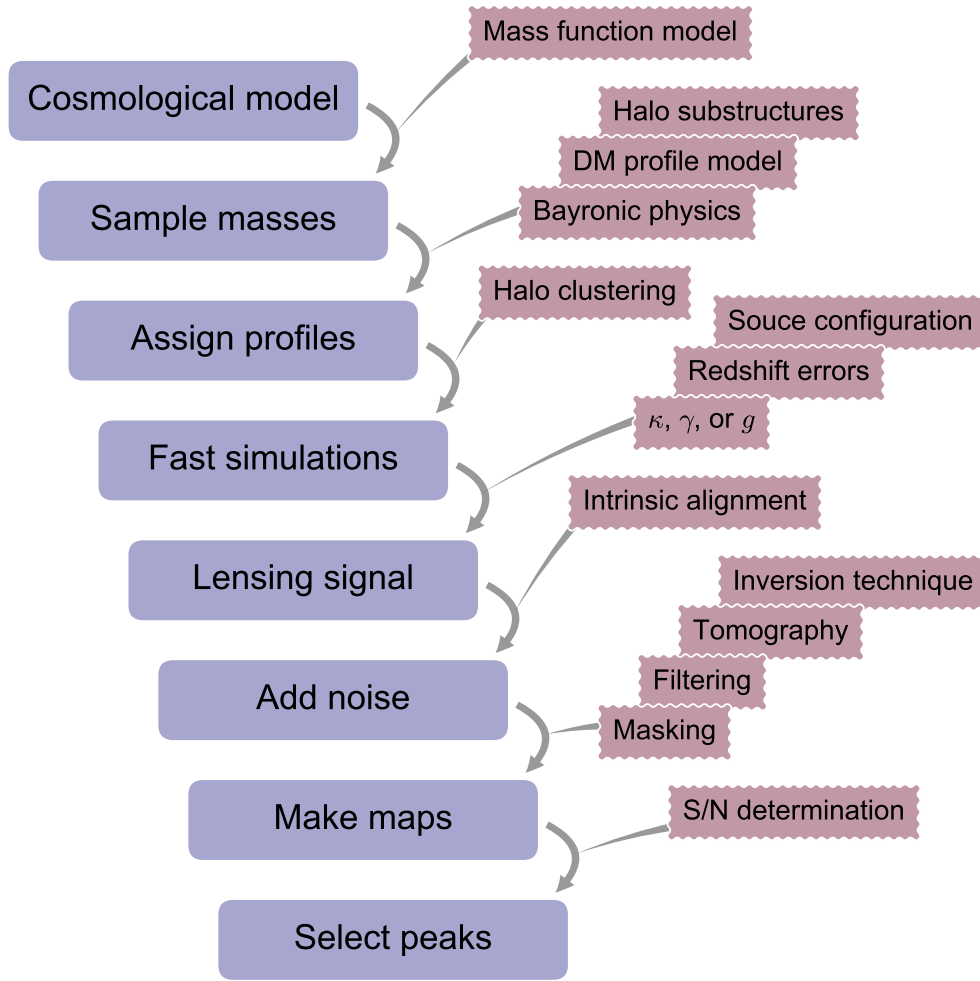
FIGURE 5.10: An indicative diagram listing possible extensions for our model. From each blue rectangle to another, red stamps are more physics to consider, observational effects, or processing techniques which would improve the understanding or the precision of WL peak counts. Implicitly, this diagram implies the forward nature of our model.

As a summary, our model is capable to take into account a wide range of potential systematic sources: astrophysical ones (halo modelling, baryons, IA), observational ones (masking, redshift errors), or the ones from data processing (κ-γ inversion, inpainting). For the reason of time, not all topics are addressed in this thesis. The settings chosen will be detailed in each analysis.

5.5 Sensitivity to cosmology

Before addressing a detailed study on parameter constraints in Chap. 6, I will show here a first look at the sensitivity of the FSF modelling to some cosmological parameters. Is our model sensitive to cosmology? Here, without performing constraints, some model realizations are shown by varying the matter density Ω_{m} and the matter fluctuation amplitude σ_8. The aim is to give an idea of different behaviors of the model regarding to these two most lensing-sensitive parameters.
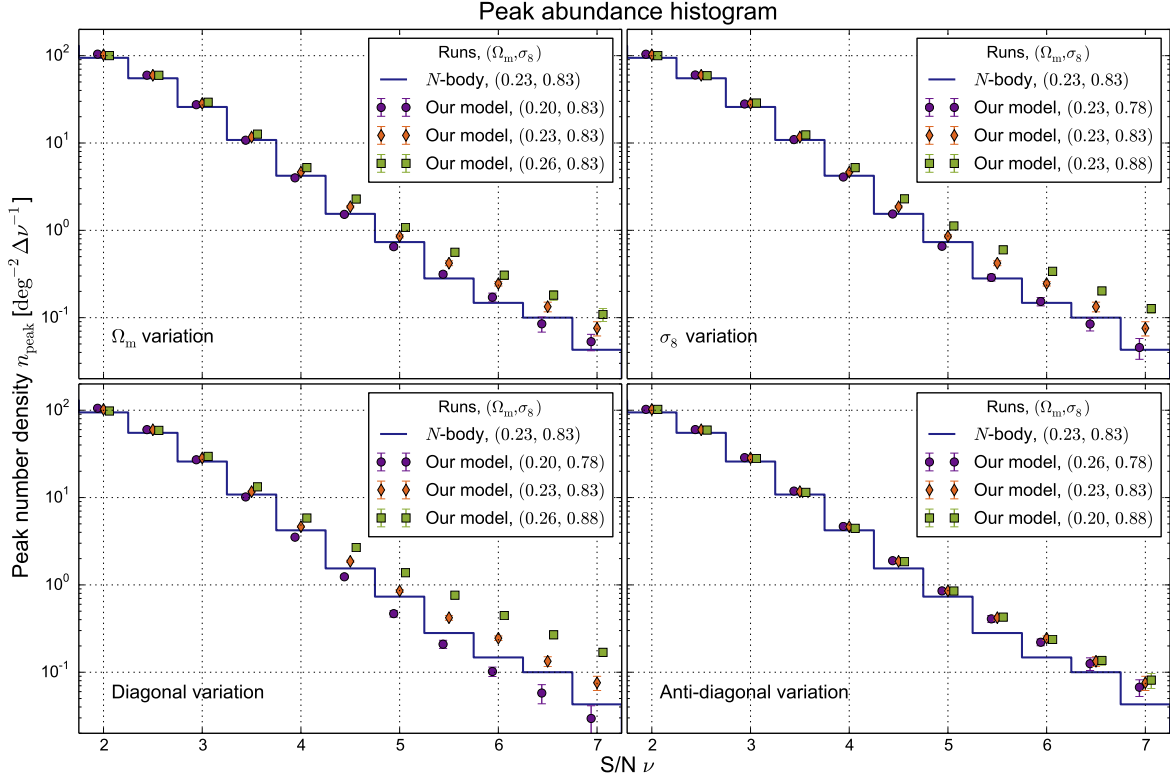
FIGURE 5.11: Peak abundance from our model for different cosmological parameter sets. The four panels display different directions of variation on the Ω_{m}-σ_8 plane. All orange diamonds are the same runs as in Fig. 5.9. This test shows that the degeneracy lines correspond approximately to the anti-diagonal direction.

On the Ω_{m}-σ_8 plane, constraints from lensing second-order statistics tend to follow curved degeneracy lines, which results in "banana-shaped" contours. Thus, four groups of parameter sets have been proposed to be compared with the input cosmology of Aardvark $(\Omega_{\mathrm{m}}^{\mathrm{in}}, \sigma_8^{\mathrm{in}}) = (0.23, 0.83)$. With $\Delta\Omega_{\mathrm{m}} = 0.03$ and $\Delta\sigma_8 = 0.05$, these four groups are successively $(\Omega_{\mathrm{m}}^{\mathrm{in}} \pm \Delta\Omega_{\mathrm{m}}, \sigma_8^{\mathrm{in}})$, $(\Omega_{\mathrm{m}}^{\mathrm{in}}, \sigma_8^{\mathrm{in}} \pm \Delta\sigma_8)$, $(\Omega_{\mathrm{m}}^{\mathrm{in}} \pm \Delta\Omega_{\mathrm{m}}, \sigma_8^{\mathrm{in}} \pm \Delta\sigma_8)$, and $(\Omega_{\mathrm{m}}^{\mathrm{in}} \pm \Delta\Omega_{\mathrm{m}}, \sigma_8^{\mathrm{in}} \mp \Delta\sigma_8)$. Thus, the first two groups vary only one of the parameters at a time, and the two others are variations along the diagonal ($45°$) and the anti-diagonal ($135°$) directions, respectively. This makes nine different parameter sets in total, including $(\Omega_{\mathrm{m}}^{\mathrm{in}}, \sigma_8^{\mathrm{in}})$.

The four panels of Fig. 5.11 show the comparison between models with different cosmologies. The variation of high-peak counts is neatly discernible when cosmology changes. Both bottom panels indicate that the degeneracy lines follow rather the anti-diagonal direction. This conforms to the a banana-shaped constraint contour suggested by other lensing observables. In addition, if the degeneracy slope of peak counts differs from the power spectrum, combining information from both would improve cosmological constraints (Dietrich & Hartlap 2010).

The blue lines in Fig. 5.11 are peak counts from Case 1, which is considered as the "truth". Under this consideration, one may estimates that the bias from our model is on the order of 0.03 for Ω_{m}, or equivalent to ~ 0.05 for σ_8. However, one should remind that this "truth" is only computed on the small field while all nine model runs are on the large one. Therefore, the "truth" is only indicative. Nevertheless, the model is not perfect and more studies are needed in order to understand the sources of biases.

5.6 Comparison with an analytical model

5.6.1 Description of the FSL model

The last part of this chapter focuses on the model from Fan, Shan, & Liu (2010, hereafter FSL model). Based on an analytical approach, it computes directly the values of the peak function for various ν from the peak theory. Consider a smoothed Gaussian random field $K_N = K + N$ fluctuating around its mean value K with a zero-mean random part N. In the context of WL peaks, N characterizes galaxy shape noise and K is just the projected mass of foreground halos. Given a position, the joint probability of values of K_N and its derivatives is well known. Now, if a point is a local maximum, the first derivatives should be zero and both eigenvalues of the second derivative matrix should be negative. These two conditions lead to determine a peak probability function which depends on the values and the derivatives of the foreground.

For the FSL model, all is reasoned in terms of probability. The requirement of the FSL model is also a mass function model and a halo profile. The mass function characterizes the halo population as a probability function of mass and redshift. In other words, it may be interpreted as, given a line of sight, the probability of crossing a halo of mass M at redshift z. Now, if one ensures that the halo does not overlap with others on the line of sight and if the profile is known, then one knows with which probability the line of sight pierces the halo at a particular angular position. Knowing the relative position to the halo, one also obtains the smoothed convergence and the derivatives. And by combining the probabilities of the halo population, the relative position, the random fluctuation, and the condition of being peaks, one reaches eventually the peak function.

The basic assumption of the FSL model is that halos do not overlap on the 2D sky. Actually, this assumption is not necessary because the probability of having two halos on the line of sight can be characterized by a term proportional to the product of two mass functions. However, this computation is very expensive, and since the non-overlapping hypothesis has been proven to be a good approximation (Yang et al. 2011), the following derivation only focuses on the first order of the FSL model.

5.6.2 Formalisms

Joint PDF from a Gaussian random field

Consider a Gaussian random field n. Let n be smoothed by a kernel function W to become $N \equiv n * W$, assumed to have zero mean everywhere. The smoothed field N is characterized by its moments σ_i for $i = 0, 1, 2, \ldots$, defined as (Van Waerbeke 2000)

$$\sigma_i^2 \equiv \int \frac{\mathrm{d}^2\boldsymbol{\ell}}{(2\pi)^2} \, \ell^{2i} |\widetilde{N}(\ell)|^2, \tag{5.8}$$

where $\widetilde{N}$ is the Fourier transform of N. Locally on the field, peak selection depends on the value and the first and the second derivatives of the considered position. In a 2D space, these are the value N, two first derivative components $N_{,1}$ and $N_{,2}$, and three effective terms from second derivatives $N_{,11}$, $N_{,22}$, and $N_{,12}$. Denoting $\boldsymbol{x} = (N, N_{,1}, N_{,2}, N_{,11}, N_{,22}, N_{,12})$ as the

considered random vector, the six-dimensional joint Gaussian PDF is

$$P(N, N_{,1}, N_{,2}, N_{,11}, N_{,22}, N_{,12}) = \frac{1}{\sqrt{(2\pi)^6 \det |\boldsymbol{C}|}} \exp\left[-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\right], \tag{5.9}$$

where the covariance matrix $\boldsymbol{C}$ is (Van Waerbeke 2000)

$$\boldsymbol{C} = \begin{pmatrix} \sigma_0^2 & 0 & 0 & -\sigma_1^2/2 & -\sigma_1^2/2 & 0 \\ 0 & \sigma_1^2/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_1^2/2 & 0 & 0 & 0 \\ -\sigma_1^2/2 & 0 & 0 & 3\sigma_2^2/8 & \sigma_2^2/8 & 0 \\ -\sigma_1^2/2 & 0 & 0 & \sigma_2^2/8 & 3\sigma_2^2/8 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_2^2/8 \end{pmatrix}. \tag{5.10}$$

After simple calculations, Eq. (5.9) becomes

$$P\left(\widetilde{N}, \widetilde{N}_{,1}, \widetilde{N}_{,2}, \widetilde{N}_{,11}, \widetilde{N}_{,22}, \widetilde{N}_{,12}\right)$$

$$= \frac{2}{\pi^3\sqrt{1-\gamma_*^2}} \exp\left[-\frac{\widetilde{N}^2 + 2\gamma_*\widetilde{N}\left(\widetilde{N}_{,11} + \widetilde{N}_{,22}\right) + \left(\widetilde{N}_{,11} + \widetilde{N}_{,22}\right)^2}{2\left(1-\gamma_*^2\right)}\right]$$

$$\times \exp\left[-\left(\widetilde{N}_{,11} - \widetilde{N}_{,22}\right)^2 - 4\widetilde{N}_{,12}{}^2 - \widetilde{N}_{,1}{}^2 - \widetilde{N}_{,2}{}^2\right], \tag{5.11}$$

where $\widetilde{N} \equiv N/\sigma_0$, $\widetilde{N}_{,i} \equiv N_{,i}/\sigma_1$, and $\widetilde{N}_{,ij} \equiv N_{,ij}/\sigma_2$ for $i, j \in \{1, 2\}$, and $\gamma_* \equiv \sigma_1^2/\sigma_0\sigma_2$.

Peak density as a function of the foreground

When the foreground clusters are present, $K \neq 0$. One should not count peaks from N but from $K_N = K + N$. To write down the joint probability for quantities related to K_N, one only needs to replace N by $K_N - K$ in Eq. (5.11). Before doing so, it is useful to apply some changes of variable as follows. First, with the same logic as before, I use dimensionless quantities:

$$\widetilde{K} \equiv K/\sigma_0, \quad \widetilde{K}_{,i} \equiv \partial_i K/\sigma_1, \quad \widetilde{K}_{,ij} \equiv \partial_i\partial_j K/\sigma_2,$$
$$\widetilde{K}_N \equiv K_N/\sigma_0, \quad \widetilde{K}_{N,i} \equiv \partial_i K_N/\sigma_1, \quad \widetilde{K}_{N,ij} \equiv \partial_i\partial_j K_N/\sigma_2. \tag{5.12}$$

Further, since the zero-order moment σ_0 is nothing but $\sigma_{\rm noise}$, K_N/σ_0 is the S/N. Therefore, one may set $\nu = \widetilde{K}_N$ to simplify the notation. Then, consider the second derivatives of $\widetilde{K}_N$. The matrix $-\widetilde{K}_{N,ij}$ is symmetric, so it has two real eigenvalues, say λ_{N1} and λ_{N2} with $\lambda_{N1} \geq \lambda_{N2}$. By defining

$$x_N \equiv \lambda_{N1} + \lambda_{N2} \quad \text{and} \quad e_N \equiv \frac{\lambda_{N1} - \lambda_{N2}}{2(\lambda_{N1} + \lambda_{N2})}, \tag{5.13}$$

the second derivatives can be parametrized by these quantities as

$$\widetilde{K}_{N,11} = -\frac{x_N}{2}\left(1 + 2e_N\cos 2\theta_N\right), \tag{5.14}$$

$$\widetilde{K}_{N,22} = -\frac{x_N}{2}\left(1 - 2e_N\cos 2\theta_N\right), \tag{5.15}$$

$$\widetilde{K}_{N,12} = -x_N e_N \sin 2\theta_N. \tag{5.16}$$

where θ_N is a mixing angle in $[0, \pi]$. After the change of variables, taking into account the Jacobian matrix, the probability from Eq. (5.11) becomes (see Bardeen et al. 1986, Bond & Efstathiou 1987, and Fan et al. 2010)

$$P\left(\nu, x_N, e_N, \theta_N, \widetilde{K}_{N,1}, \widetilde{K}_{N,2}\right)$$

$$= \frac{4x_N^2 e_N}{\pi^3 \sqrt{1-\gamma_*^2}} \, \exp\left[-\frac{1}{2}\left(\nu - \widetilde{K}\right)^2 - \frac{\left(x_N + \widetilde{K}_{,11} + \widetilde{K}_{,22} - \gamma_*\left(\nu - \widetilde{K}\right)\right)^2}{2\left(1-\gamma_*^2\right)}\right]$$

$$\times \exp\left[-4x_N^2 e_N^2 - 4x_N e_N \cos 2\theta_N \left(\widetilde{K}_{,11} - \widetilde{K}_{,22}\right) - \left(\widetilde{K}_{,11} - \widetilde{K}_{,22}\right)^2\right]$$

$$\times \exp\left[-8x_N e_N \sin 2\theta_N \widetilde{K}_{,12} - 4\widetilde{K}_{,12}^2\right]$$

$$\times \exp\left[-\left(\widetilde{K}_{N,1} - \widetilde{K}_{,1}\right)^2 - \left(\widetilde{K}_{N,2} - \widetilde{K}_{,2}\right)^2\right]. \tag{5.17}$$

Here, Eq. (5.17) is the probability for general cases. Two conditions need to be added for peak selection: the vanishing of the first derivatives and negative eigenvalues of $\widetilde{K}_{N,ij}$. The first one is ensured by adding $\delta(\widetilde{K}_{N,1})\delta(\widetilde{K}_{N,2})$, where δ stands for the Dirac delta function, and the second is translated by the condition $\lambda_{N1} \geq \lambda_{N2} \geq 0$, or more precisely $\Theta(x_N)\Theta(e_N)\Theta(1 - 2e_N)\Theta(\theta_N)\Theta(\pi - \theta_N)$ with Θ being the Heaviside step function. Moreover, a surface term $(\sigma_2/\sigma_1)^2(x_N^2/4)(1 - 4e_N^2)$ should be introduced to transform Eq. (5.17) from a one-point probability into a normalized quantity, $n_{\text{peak}}^{\text{los}}$, such that the dimension is equal to a surface density. The label los stands for line of sight [3]. At the end of the day, $n_{\text{peak}}^{\text{los}}$ is linked to "foreground quantities" $\widetilde{K}$, $\widetilde{K}_{,i}$, and $\widetilde{K}_{,ij}$ via

$$n_{\text{peak}}^{\text{los}}\left(\nu \Big| \widetilde{K}, \widetilde{K}_{,1}, \widetilde{K}_{,2}, \widetilde{K}_{,11}, \widetilde{K}_{,22}, \widetilde{K}_{,12}\right)$$

$$= \int_0^{+\infty} \mathrm{d}x_N \int_0^{1/2} \mathrm{d}e_N \int_0^\pi \mathrm{d}\theta_N \int_\mathbb{R} \mathrm{d}\widetilde{K}_{N,1} \, \delta(\widetilde{K}_{N,1}) \int_\mathbb{R} \mathrm{d}\widetilde{K}_{N,2} \, \delta(\widetilde{K}_{N,2})$$

$$\times P\left(\nu, x_N, e_N, \theta_N, \widetilde{K}_{N,1}, \widetilde{K}_{N,2}\right) \cdot (\sigma_2/\sigma_1)^2 \left(x_N^2/4\right)\left(1 - 4e_N^2\right), \tag{5.18}$$

which I write explicitly as

$$n_{\text{peak}}^{\text{los}}\left(\nu \Big| \widetilde{K}, \widetilde{K}_{,1}, \widetilde{K}_{,2}, \widetilde{K}_{,11}, \widetilde{K}_{,22}, \widetilde{K}_{,12}\right)$$

$$= \frac{1}{4\pi^2 \theta_*^2 \sqrt{1-\gamma_*^2}} \, \exp\left[-\frac{1}{2}\left(\nu - \widetilde{K}\right)^2 - \widetilde{K}_{,1}^2 - \widetilde{K}_{,2}^2\right]$$

$$\times \int_0^{+\infty} \mathrm{d}x_N \, \exp\left[-\frac{\left(x_N + \widetilde{K}_{,11} + \widetilde{K}_{,22} - \gamma_*\left(\nu - \widetilde{K}\right)\right)^2}{2\left(1-\gamma_*^2\right)}\right]$$

$$\times x_N^4 \exp\left[-\left(\widetilde{K}_{,11} - \widetilde{K}_{,22}\right)^2 - 4\widetilde{K}_{,12}^2\right] \cdot \int_0^{1/2} \mathrm{d}e_N \, \exp\left[-4x_N^2 e_N^2\right] \cdot 8e_N\left(1 - 4e_N^2\right)$$

$$\times \int_0^\pi \frac{\mathrm{d}\theta_N}{\pi} \, \exp\left[-4x_N e_N \cos 2\theta_N \left(\widetilde{K}_{,11} - \widetilde{K}_{,22}\right) - 8x_N e_N \sin 2\theta_N \widetilde{K}_{,12}\right], \tag{5.19}$$

where $\theta_*^2 \equiv 2\sigma_1^2/\sigma_2^2$ is equivalent to a surface.

In principle, one can integrate Eq. (5.19) numerically, since the number of dimension is only three. However, as we are going to see later, the total integral also contains S/N, redshift, mass, and radius. Whatever the size of the cosmological parameter space is, this constraint requires already seven additional dimensions while performing model evaluation. In order to

---

[3] The "number density for a line of sight" does not make sense, so I avoid using this interpretation. It should rather be understood, as I said, as a probability which has a dimension of a surface density.

---

reduce the cost, Fan et al. (2010) have added some short cuts. Define $F(x_N)$ as the two last lines of Eq. (5.19), which is

$$
\begin{aligned}
F(x_N) \equiv{}& x_N^4 \exp\left[-\left(\widetilde{K}_{,11} - \widetilde{K}_{,22}\right)^2 - 4\widetilde{K}_{,12}{}^2\right] \\
&\times \int_0^{1/2} \mathrm{d}e_N \ \exp\left[-4x_N^2 e_N^2\right] \cdot 8e_N\left(1 - 4e_N^2\right) \\
&\times \int_0^\pi \frac{\mathrm{d}\theta_N}{\pi} \ \exp\left[-4x_N e_N \cos 2\theta_N\left(\widetilde{K}_{,11} - \widetilde{K}_{,22}\right) - 8x_N e_N \sin 2\theta_N \widetilde{K}_{,12}\right],
\end{aligned}
\tag{5.20}
$$

If no foreground is present, this function has an analytical solution which is $F(x_N) = x_N^2 - 1 + \exp(-x_N^2)$ (Bond & Efstathiou 1987). In the current scenario, if one considers that the foreground contains only one spherical halo on the line of sight, then Eq. (5.20) can be simplified. The spherical symmetry allows one to choose a new coordinate system such that $\widetilde{K}_{,2}$ and $\widetilde{K}_{,12}$ vanish. Therefore, $F(x_N)$ becomes $F(x_N, \widetilde{K}_{,11} - \widetilde{K}_{,22})$. Inspired by the exact solution in the case without foreground, Fan et al. (2010) have proposed an approximate form for F:

$$
F\left(x_N, \widetilde{K}_\mathrm{diff}\right) = x_N^2 - 1 + \exp\left(-x_N^2\right) + \widetilde{K}_\mathrm{diff}{}^2\left[\exp\left(-\frac{x_N^3}{g(\widetilde{K}_\mathrm{diff})}\right) - 1\right],
\tag{5.21}
$$

where $\widetilde{K}_\mathrm{diff} \equiv \widetilde{K}_{,11} - \widetilde{K}_{,22}$ and g is a function of $\widetilde{K}_\mathrm{diff}$ whose values are to be determined by fitting. In this way, the integration over e_N and θ_N only needs to be evaluated once. Let $\widetilde{K}_\mathrm{sum} \equiv \widetilde{K}_{,11} + \widetilde{K}_{,22}$. In the coordinate system where $\widetilde{K}_{,2}$ and $\widetilde{K}_{,12}$ vanish, one can rewrite Eq. (5.19) as

$$
\begin{aligned}
n_\mathrm{peak}^\mathrm{los}\left(\nu\middle|\widetilde{K}, \widetilde{K}_{,1}, \widetilde{K}_\mathrm{sum}, \widetilde{K}_\mathrm{diff}\right) ={}& \frac{1}{4\pi^2\theta_*^2\sqrt{1 - \gamma_*^2}} \ \exp\left[-\frac{1}{2}\left(\nu - \widetilde{K}\right)^2 - \widetilde{K}_{,1}{}^2\right] \\
&\times \int_0^{+\infty} \mathrm{d}x_N \ F\left(x_N, \widetilde{K}_\mathrm{diff}\right)\exp\left[-\frac{\left(x_N + \widetilde{K}_\mathrm{sum} - \gamma_*(\nu - \widetilde{K})\right)^2}{2\left(1 - \gamma_*^2\right)}\right],
\end{aligned}
\tag{5.22}
$$

where $F(x_N)$ is of course given by Eq. (5.21).

Densities from two regimes: halo-covered and non-covered areas

Now we need to compute, for a halo of mass M at redshift z, the values of $\widetilde{K}$ and its derivatives. Fan et al. (2010) have considered NFW profiles (Eq. (3.86)) and have set $\kappa = \kappa_\mathrm{proj}$ (Eqs. (4.62), (4.63), and (5.7)) without correction. The normalized smoothed convergence $\widetilde{K}$ is obtained via

$$
\widetilde{K}(\boldsymbol{\theta}) = \sigma_0^{-1} \int \mathrm{d}^2\boldsymbol{\theta}' \ W(\boldsymbol{\theta} - \boldsymbol{\theta}')\kappa(\boldsymbol{\theta}').
\tag{5.23}
$$

And from Eq. (5.23) we derive easily

$$
\widetilde{K}_{,i}(\boldsymbol{\theta}) = \sigma_1^{-1} \int \mathrm{d}^2\boldsymbol{\theta}' \ \partial_i W(\boldsymbol{\theta} - \boldsymbol{\theta}')\kappa(\boldsymbol{\theta}'),
\tag{5.24}
$$

$$
\widetilde{K}_{,ij}(\boldsymbol{\theta}) = \sigma_2^{-1} \int \mathrm{d}^2\boldsymbol{\theta}' \ \partial_i\partial_j W(\boldsymbol{\theta} - \boldsymbol{\theta}')\kappa(\boldsymbol{\theta}').
\tag{5.25}
$$

If one considers the kernel from Eq. (5.3), which is the same as in Fan et al. (2010), these quantities become

$$\widetilde{K}(\theta_1, 0) = \int d^2\boldsymbol{\theta}' \; \frac{1}{\pi\theta_G^2\sigma_0} \exp\left[-\frac{(\theta_1 - \theta_1')^2 + (-\theta_2')^2}{\theta_G^2}\right] \kappa(\boldsymbol{\theta}'), \tag{5.26}$$

$$\widetilde{K}_{,1}(\theta_1, 0) = \int d^2\boldsymbol{\theta}' \; -\frac{2}{\theta_G^2}(\theta_1 - \theta_1') \cdot \frac{1}{\pi\theta_G^2\sigma_1} \exp\left[-\frac{(\theta_1 - \theta_1')^2 + (-\theta_2')^2}{\theta_G^2}\right] \kappa(\boldsymbol{\theta}'), \tag{5.27}$$

$$\widetilde{K}_{\text{sum}}(\theta_1, 0) = \int d^2\boldsymbol{\theta}' \; \frac{4}{\theta_G^2}\left[\frac{(\theta_1 - \theta_1')^2 + (-\theta_2')^2}{\theta_G^2} - 1\right] \cdot \frac{1}{\pi\theta_G^2\sigma_2} \exp\left[-\frac{(\theta_1 - \theta_1')^2 + (-\theta_2')^2}{\theta_G^2}\right] \kappa(\boldsymbol{\theta}'), \tag{5.28}$$

$$\widetilde{K}_{\text{diff}}(\theta_1, 0) = \int d^2\boldsymbol{\theta}' \; \frac{4}{\theta_G^2}\left[\frac{(\theta_1 - \theta_1')^2 - (-\theta_2')^2}{\theta_G^2}\right] \cdot \frac{1}{\pi\theta_G^2\sigma_2} \exp\left[-\frac{(\theta_1 - \theta_1')^2 + (-\theta_2')^2}{\theta_G^2}\right] \kappa(\boldsymbol{\theta}'). \tag{5.29}$$

Only $\theta_2 = 0$ is taken into account because for a spherical profile, this condition leads to $\widetilde{K}_{,2} = \widetilde{K}_{,12} = 0$. Readers should be reminded that κ in Eqs. (5.26), (5.27), (5.28), and (5.29) depends on halo redshift and mass. Therefore, $\widetilde{K}$, $\widetilde{K}_{,1}$, $\widetilde{K}_{\text{sum}}$, and $\widetilde{K}_{\text{diff}}$ are implicit functions of z, M, and θ_1. Finally, for a field of view of size $d^2\Omega$, the total number of peaks from halos is given by

$$n_{\text{peak}}^{\text{halo}}(\nu)d^2\Omega = \int dz \; \frac{dV(z)}{dz} \int d\log M \; \frac{dn(z, <M)}{d\log M} \int_0^{\theta_{\text{vir}}} d\theta_1 \; 2\pi\theta_1$$
$$\times n_{\text{peak}}^{\text{los}}\left(\nu \Big| \widetilde{K}(z, M, \theta_1), \widetilde{K}_{,1}(z, M, \theta_1), \widetilde{K}_{\text{sum}}(z, M, \theta_1), \widetilde{K}_{\text{diff}}(z, M, \theta_1)\right), \tag{5.30}$$

where

$$dV(z) = dw \cdot w^2(z)d^2\Omega = \frac{D_H \cdot w^2(z)}{E(z)} \; dz d^2\Omega \tag{5.31}$$

is the comoving volume contained in a slice of redshift $[z, z + dz[$ and $dn(z, <M)/d\log M$ is the halo volume density with log mass contained between $\log M$ and $\log M + d\log M$ (defined as Eq. (3.85)). The term $E(z)$ is given by Eq. (2.19).

Equation (5.30) is straightforward to interpret. The integration over z yields the light-cone volume for $d^2\Omega$; the integration over $\log M$ gives the number of halos in the lightcone; the integration over θ_1 shows the total covered surface by these halos; and finally $n_{\text{peak}}^{\text{halo}}$ represents the probability by unity of surface that a peak at level ν would be observed. However, this is not the end of the story. Equation (5.30) only takes into account those lines of sight where a halo is present. That is the reason for the label `halo`. From the sky area not covered by halos, peaks can still occur from random fluctuations. This foreground-free case is described by Eq. (5.22) by setting all K-related quantities to zero, leading to

$$n_{\text{peak}}^{\text{los}}\left(\nu \Big| \widetilde{K} = \widetilde{K}_{,1} = \widetilde{K}_{\text{sum}} = \widetilde{K}_{\text{diff}} = 0\right)$$
$$= \frac{\exp(-\nu^2/2)}{4\pi^2\theta_*^2\sqrt{1 - \gamma_*^2}} \int_0^{+\infty} dx_N \; \left(x_N^2 - 1 + \exp(-x_N^2)\right) \exp\left[-\frac{(x_N - \gamma_*\nu)^2}{2(1 - \gamma_*^2)}\right], \tag{5.32}$$

and the peak density occurs from pure random fluctuation $n_{\text{peak}}^{\text{ran}}$ reads

$$n_{\text{peak}}^{\text{ran}}(\nu)d^2\Omega = \left(d^2\Omega - \int dz \; \frac{dV(z)}{dz} \int d\log M \; \frac{dn(z, <M)}{d\log M} \int_0^{\theta_{\text{vir}}} d\theta_1 \; 2\pi\theta_1\right)$$
$$\times n_{\text{peak}}^{\text{los}}\left(\nu \Big| \widetilde{K} = \widetilde{K}_{,1} = \widetilde{K}_{\text{sum}} = \widetilde{K}_{\text{diff}} = 0\right). \tag{5.33}$$
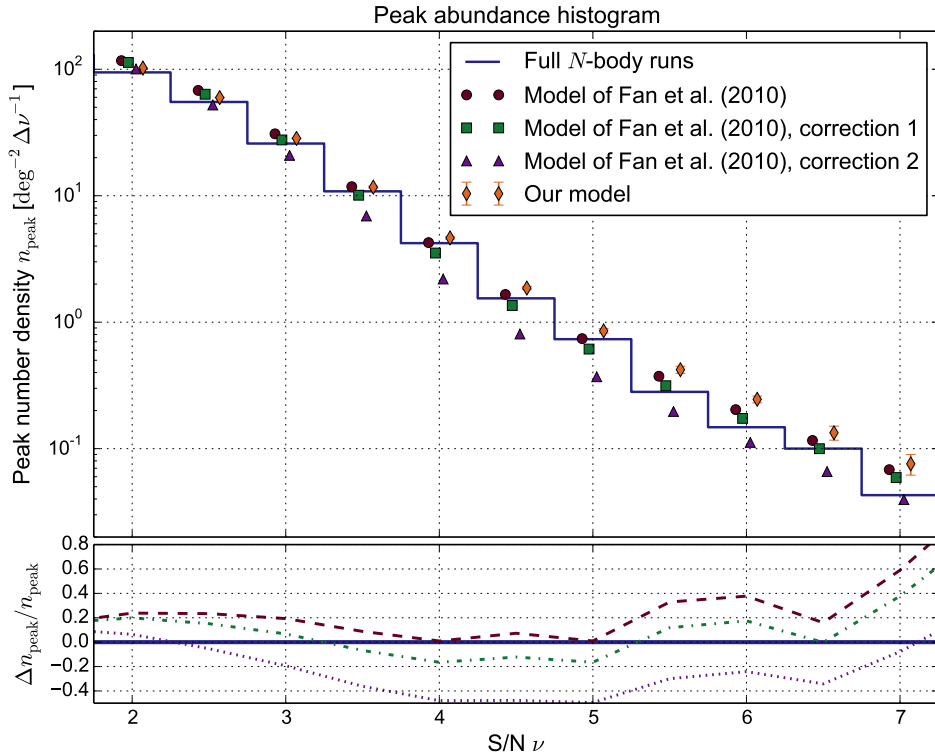
FIGURE 5.12: Comparison with an analytical model. On the upper panel, we show peak histograms from N-body runs (blue solid lines, Case 1 in Sect. 5.3.2), our model (orange diamonds, Case 4 in Sect. 5.3.2), the FSL model (dark red circles), and two corrections (green squares and purple triangles) for the latter (see Sect. 5.6.3). On the lower panel, the deviations of all cases with regard to N-body runs are shown. The orange band delineates the upper and lower limits of the error bars of our model, whereas for the other three cases, the lines indicate the mean. Results from N-body runs are obtained from the small field, while those from our model are performed on the large field.

As a result, combining Eqs. (5.30) and (5.33), the peak number density from the FSL model is

$$n_{\mathrm{peak}}(\nu) = n_{\mathrm{peak}}^{\mathrm{halo}}(\nu) + n_{\mathrm{peak}}^{\mathrm{ran}}(\nu). \tag{5.34}$$

5.6.3   Results

Figure 5.12 shows the comparison between N-body runs, the FSL model, and our model with the same input cosmology. Both models use the same noise level and smoothing scale $\theta_{\mathrm{G}} = 1$ arcmin. The lower panel shows the relative difference of each model compared to the N-body runs. It is worth highlighting that the histogram of N-body runs (blue lines) is extracted from a patch of 54 deg^2. The cosmic-variance effect might be important. Nevertheless, both models show an overall good agreement with the Aardvark N-body simulations (dark red circles and orange diamonds; the two other cases will be discussed later).

At first glance, Fig. 5.12 seems to suggest that the FSL model performs better. However, several problems can be revealed if by looking closer. As I mentioned above, the FSL model is based on the assumption that halos do not overlap in projection. The direct consequence of this is that a cutoff at low redshift needs to be introduced. Otherwise, whatever the physical size of the halo is, the angular size would converge to the half of the sky (2π rad^2) when $z \to 0$. The total halo-covered area would be much larger than the field of view in this case.

Taking into account that low-redshift halos have poor lensing efficiency, this probably does not create any significant bias.

The other restriction related to the total halo-covered area is the lower mass limit. If $M_{\min}$ is low, then halos will be numerous. Of course, most of them will be small and have a reduced angular size, but it seems that the decreasing rate of the angular size is dominated by the increasing rate of the population as $M_{\min}$ decreases. As a result, by $M_{\min} \lesssim 10^{13.3}\ M_\odot/h$, the halo-covered area will exceed the total field of view and break down the model, since it assumes no overlap in projection between halos. For the FSL model in Fig. 5.12, $M_{\min}$ is set to $10^{13.6}\ M_\odot/h$. However, as indicated by Fig. 5.8, this value is already large enough to create some deficits on middle-peak counts. Recall that our model has $M_{\min} = 10^{12}\ M_\odot/h$. By increasing $M_{\min}$, we expect that the orange diamonds would move closer to the blue solid lines; this however does not imply a better modelling! As a conclusion, the FSL model would have faced the same augmentation if it manages to go lower than $M_{\min} = 10^{13.3}\ M_\odot/h$. A more detailed analysis should be done in the future, especially for comparing with a larger N-body reference field.

Revising the κ_{proj} formalism

Even though high-peak counts are underestimated by the FSL model, this result presents already an excess. What is the origin of these overcounts? Readers may have noticed that, contrary to our model, the FSL model assumes $\kappa = \kappa_{\mathrm{proj}}$. Not subtracting anything from κ_{proj} leads systematically to an overestimation of peak counts. The space which is not occupied by halos is not empty (see Sect. 4.2.1)! The projected mass accounted in Eqs. (5.30) and (5.33) are only carried out over $[M_{\min}, M_{\max}[$, which is somewhat between 10% and 40% of the total mass. If this additional mass is not extracted, peak counts can be largely influenced.

Below I propose a correction for the FSL model. Let us rewrite the matter density as

$$\rho(f_K(w)\boldsymbol{\theta}, w) = \sum_{\mathrm{halos}} \rho_{\mathrm{halo}}(f_K(w)\boldsymbol{\theta}, w) + (1 - \lambda(w))\bar{\rho}(w), \tag{5.35}$$

where ρ_{halo} is halo density profiles and

$$\lambda(w(z)) \equiv \frac{1}{\rho_{\mathrm{crit}}\Omega_{\mathrm{m}}} \int_{M_{\min}}^{M_{\max}} \mathrm{d}\log M\ \frac{\mathrm{d}n(z, <M)}{\mathrm{d}\log M} \cdot M, \tag{5.36}$$

The quantity $\lambda(w)$ is the proportion of the mass falling in $[M_{\min}, M_{\max}[$ at the epoch corresponding to w from the observer [♯4]. With the factor $1 - \lambda(w)$, the new mean density is now the correct one as $\bar{\rho}(w) = \rho_{\mathrm{crit}}\Omega_{\mathrm{m}}(1 + z(w))^3$. The resulting density contrast is

$$\delta(f_K(w)\boldsymbol{\theta}, w) = \sum_{\mathrm{halos}} \frac{\rho_{\mathrm{halo}}(f_K(w)\boldsymbol{\theta}, w)}{\bar{\rho}(w)} - \lambda(w). \tag{5.37}$$

Now, if we insert Eq. (5.37) into Eq. (4.36), the definition of κ, we find

$$\kappa(\boldsymbol{\theta}, w) = \kappa_{\mathrm{proj}}(\boldsymbol{\theta}, w) - \kappa_1(w) \tag{5.38}$$

[♯4]No scaling factor should appear in Eq. (5.36). The density that the mass function gives is per comoving volume. The matter density, usually denoted as $\rho(z)$, is per physical volume at z. In terms of comoving volume, it is per physical volume at $z = 0$, which is $\rho_{\mathrm{crit}}\Omega_{\mathrm{m}}$.

with $\kappa_{\mathrm{proj}}(\boldsymbol{\theta}, w)$ given by Eq. (5.7) and

$$\kappa_1(w) \equiv \frac{3H_0^2\Omega_{\mathrm{m}}}{2\mathrm{c}^2} \int_0^w \mathrm{d}w' \, \frac{f_K(w - w')f_K(w')}{f_K(w)} \frac{\lambda(w')}{a(w')}. \tag{5.39}$$

For z_{s} fixed at 1, κ_1 is a constant. The numerical result implies that $\kappa_1 = 0.0034$ for $M_{\mathrm{min}} = 10^{13.6} \, M_\odot/h$ in the Aardvark cosmology, which is 0.14σ for a Gaussian smoothing of $\theta_{\mathrm{G}} = 1$ arcmin. In Fig. 5.12, the result taking κ_1 into account is labelled as `correction1`. Readers can see that the κ_1-correction improves neatly the model. The difference between both deviations from N-body runs (dark red dashed and green dash-dotted lines from the lower panel) tends to zero as S/N decreases. This conforms to Eq. (5.34) since $n_{\mathrm{peak}}^{\mathrm{ran}}$, which does not change, is more dominant in this regime. Computing the κ_1-correction for $M_{\mathrm{min}} = 10^{12} \, M_\odot/h$, the result is shown as `correction2` in Fig. 5.12. Of course, κ_{proj} from $M \in [10^{12}, 10^{13.6}] \, M_\odot/h$ is not taken into account in this case. However, on one hand, the comparison between Corrections 1 and 2 indicates the variation of κ_1 that one expects in each bin, showing that ignoring this difference is not appropriate; on the other hand, the comparison between Correction 2 and our model reveals the order of magnitude of the contribution from κ_{proj} for $M \in [10^{12}, 10^{13.6}] \, M_\odot/h$, which suggests that $M_{\mathrm{min}} = 10^{13.6} \, M_\odot/h$ is not pertinent.

A similar examination has been done for our model. For a setting conforming to Table 5.1, we have found $\kappa_1 = 0.013$, and $|\kappa_1 - \overline{\kappa}_{\mathrm{proj}}| \lesssim 10^{-4}$ for all realizations from our model. This displays an excellent agreement between subtracting $\overline{\kappa}_{\mathrm{proj}}$ and the $\lambda(w)$ modelling. The zero-mean convergence is therefore a good approximation. Nevertheless, for our model, applying Eq. (5.39) to the computation of the projected mass is feasible. For example, a table of $\lambda(w)$ can be precomputed for a wide range of w, then instead of calculating κ_{proj} for each galaxy and subtracting the mean after the fact, one may take directly $\kappa_{\mathrm{proj}} - \kappa_1$ as the lensing signal. This improvement will also solve the problem that $\overline{\kappa}_{\mathrm{proj}}$ depends on the cutoff size of the truncated profiles. The same improvement technique is also applicable on the FSL model.

Regarding the complexity of the realistic conditions, the FSL model seems to be limited. Modelling non-Gaussian features from photo-z errors, IA, and non-uniformly-distributed galaxies with a simple Gaussian random field would be a great challenge. Meanwhile, on the road to parameter constraints, estimating the covariance matrix would be impossible without the help of N-body simulations for analytical models like this one. This is however not the case of our model.

Summary

In this chapter, I have introduced a fast stochastic forward model to predict lensing peak counts. It performs fast simulations, which substitute the N-body physical process with a stochastic one. The advantages of the new model can be summarized to three characteristics: fast, flexible, and full PDF information.

Adopting progressively different steps of the new model, intermediate cases permit to test two major hypotheses that the model makes: neglecting unbound matter and breaking halo correlation. In spite of some biases, the model agrees well with N-body simulations. Most importantly, it has been shown that the cosmological sensitivity is larger than the biases of the model.

The origin of some systematic biases are known, including halo correlation and halo concentration. Thanks to the flexibility of the new model, many options of improvement are

possible. However, some other biases are not entirely understood. To study this, a larger N-body simulation set and more realistic probe conditions would be required for the future.

The new model has been compared to the FSL analytical model. The deviation of both from N-body runs are similar. I have also proposed a correction which improves the FSL model. However, even with the correction, the potential of the FSL model is limited by its basic assumption: non-overlapping halos in projection.

In the next chapter, I am going to demonstrate a tremendous asset of our fast stochastic model: the availability of the full PDF information, and how this can provide various constraint strategies for cosmological parameters.

Chapter VI

Parameter constraint strategies

Overview

As the link between cosmological models and peak-count observables is established, we are now able to perform model selection. Comparing cosmological models or constraining parameters requires sizing the scatter and uncertainty, and this can be done by exploring the full PDF information provided by our model. In this chapter, I am going to propose several constraining techniques. After introducing the methodology, the impact from the true covariance on the Gaussian likelihood is studied. Various alternatives to the Gaussian likelihood are then presented: the copula likelihood, the true likelihood, and model selection using p-values. This chapter corresponds to Sects. 2, 3, 4, and 5 of Paper II.

6.1 Methodology

6.1.1 Likelihood formalism

After the sensitivity test shown in Fig. 5.11, this chapter aims to present parameter constraints derived from our peak-count model. This requires a parameter estimator which is usually the likelihood function $\mathcal{L}$. Hereafter, $\boldsymbol{x}$ denotes a data vector, $\boldsymbol{\pi}$ a parameter set, and in order to avoid confusions, P is always a probability function in the space of $\boldsymbol{x}$ and $\mathcal{P}$ the one in the parameter space.

From Bayes' theorem (Bayes 1763), the conditional probability between $\boldsymbol{x}$ and $\boldsymbol{\pi}$ is

$$\mathcal{P}(\boldsymbol{\pi}|\boldsymbol{x})P(\boldsymbol{x}) = P(\boldsymbol{x}|\boldsymbol{\pi})\mathcal{P}(\boldsymbol{\pi}). \tag{6.1}$$

Let $\boldsymbol{x}^{\mathrm{obs}}$ be the observation. By setting the Bayesian evidence to unity, $P(\boldsymbol{x} = \boldsymbol{x}^{\mathrm{obs}}) = 1$, one can rewrite Eq. (6.1) as

$$\mathcal{P}(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}}) = \mathcal{L}(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}})\mathcal{P}(\boldsymbol{\pi}), \tag{6.2}$$

where $\mathcal{L}(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}}) \equiv P(\boldsymbol{x}^{\mathrm{obs}}|\boldsymbol{\pi})$ is the *likelihood function*, $\mathcal{P}(\cdot)$ the prior, and $\mathcal{P}(\cdot|\boldsymbol{x}^{\mathrm{obs}})$ the posterior. In this chapter, I will use $\mathcal{L}(\boldsymbol{\pi})$ instead of $\mathcal{L}(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}})$ to simplify the notation.
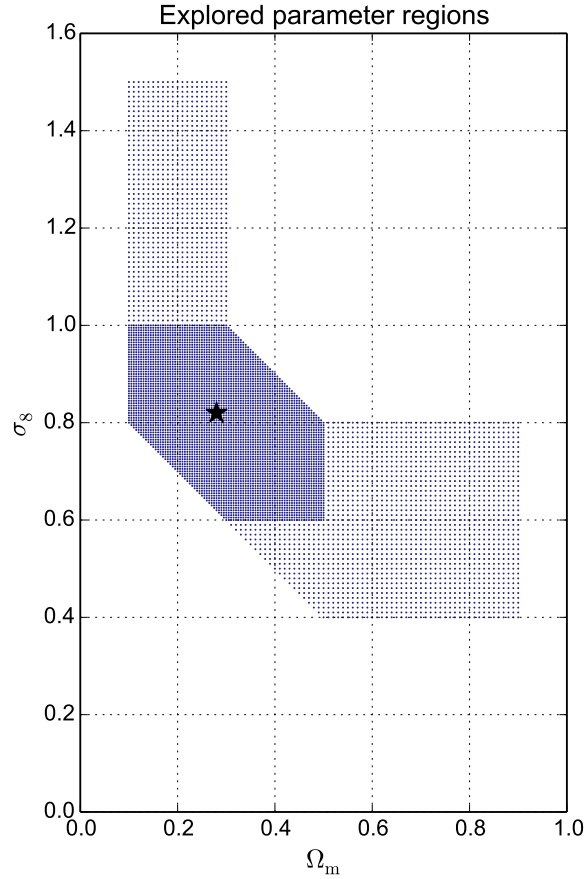
FIGURE 6.1: Location of 7821 points on which the likelihoods are evaluated. In the central condensed area, the interval between two grid points is 0.005, while in both wing zones it is 0.01. The black star displays $(\Omega_{\mathrm{m}}^{\mathrm{in}}, \sigma_8^{\mathrm{in}}) = (0.28, 0.82)$.

6.1.2 Analysis design

The study from this chapter focuses only on constraints of $\boldsymbol{\pi} = (\Omega_{\mathrm{m}}, \sigma_8)$. The other cosmological parameters are fixed, including $\Omega_{\mathrm{b}} = 0.047$, $h = 0.78$, $n_{\mathrm{s}} = 0.95$, and $w_0^{\mathrm{de}} = -1$. The dark energy density, which in this case corresponds to a cosmological constant, is noted by Ω_{Λ} and set to $1 - \Omega_{\mathrm{m}}$ to match a flat universe.

On the Ω_{m}-σ_8 plane, a region where the posterior density is potentially high is explored. A "grid-point evaluation" of the likelihood is carried out. The resolution of the grid is $\Delta\Omega_{\mathrm{m}} = \Delta\sigma_8 = 0.005$ in the center zone, and $\Delta\Omega_{\mathrm{m}} = \Delta\sigma_8 = 0.01$ in the outer area, as shown by Fig. 6.1. This results in a total of 7821 points in the parameter space to evaluate. For each parameter vector, $N = 1000$ realizations of a 25-deg^2 field are carried out. This is the data set from which statistical properties such as mean and covariance are evaluated. The setting for our peak model is similar to the description in Sect. 5.3.2 with NFW profiles, regular sources, κ peaks, Gaussian smoothing, and global significance (as sketched in Fig. 5.3). The input parameter values are detailed in Table 6.1.

Apart from the data set mentioned earlier, a single model realization under an input cosmology, $\boldsymbol{\pi}^{\mathrm{in}} = (\Omega_{\mathrm{m}}^{\mathrm{in}}, \sigma_8^{\mathrm{in}})$, is also performed. This serves as the mock observation. In other words, denoting $P(\boldsymbol{x}|\boldsymbol{\pi})$ as the PDF to obtain a data vector $\boldsymbol{x}$ from our stochastic model under the cosmology $\boldsymbol{\pi}$, the observed data vector $\boldsymbol{x}^{\mathrm{obs}}$ for this chapter is just a sample point

Parameter	Symbol	Value
Lower sampling limit	M_{min}	$10^{12}\ M_\odot/h$
Upper sampling limit	M_{max}	$10^{17}\ M_\odot/h$
Number of halo redshift bins	-	10
NFW inner slope	α	1
M-c relation amplitude	c_0	11
M-c relation power law index	β_{NFW}	0.13
Source redshift	z_{s}	1
Intrinsic ellipticity dispersion	σ_ϵ	0.4
Galaxy number density	n_{gal}	25 arcmin^{-2}
Pixel size	θ_{pix}	0.2 arcmin
Gaussian kernel size	θ_{G}	1 arcmin
Noise level in a pixel	σ_{pix}	0.283
Noise level after smoothing	σ_{noise}	0.0226
Effective field area	-	25 deg^2

TABLE 6.1: List of parameter values adopted in the study of this chapter.

drawn from $P(\cdot\,|\boldsymbol{\pi}^{\mathrm{in}})$. This study is not focusing on accuracy, but precision of the model, so that model biases and the random fluctuation related to generating a mock input can be neglected. The input parameters are set to $\Omega_{\mathrm{m}}^{\mathrm{in}} = 0.28$ and $\sigma_8^{\mathrm{in}} = 0.82$, which corresponds to a WMAP-9-like cosmology.

6.1.3 Data vector definitions

Peak-count information can be presented not only as histograms. For example, Dietrich & Hartlap (2010) have used quantities related to the inverse peak cumulative distribution function (CDF) to define observables, i.e. the data vector. Which definition can actually extract better the cosmological information? In order to answer to this question, all constraining techniques are applied on three different types of observables derived from the same data set. These three definitions of the data vector are (1) the abundance of peaks found in each S/N bin (binned peak function, noted as $\boldsymbol{x}^{\mathrm{abd}}$); (2) the S/N values at some given percentiles of the peak CDF (inverse peak CDF, noted as $\boldsymbol{x}^{\mathrm{pct}}$); and (3) similar to the second type, but with a cutoff for peaks below a certain threshold value (noted as $\boldsymbol{x}^{\mathrm{cut}}$). Mathematically, the i-th component of the two last types of observables can be denoted as x_i, thereby satisfying

$$u_i = \int_{\nu_{\mathrm{min}}}^{x_i} n_{\mathrm{peak}}(\nu)\mathrm{d}\nu, \tag{6.3}$$

where $n_{\mathrm{peak}}(\nu)$ is the peak function derived from an observation or a model realization, ν_{min} a cutoff, and u_i a specific percentile. The vector $\boldsymbol{x}^{\mathrm{pct}}$ corresponds to the case $\nu_{\mathrm{min}} = -\infty$.

The motivation for these choices is as follows. The CDF-based data vectors are free from an artificial binwidth choice, thus avoid systematics derived from binning. However, their components might be highly correlated, while diagonal terms are probably more dominant in the covariance from PDF-based data vectors. On the other hand, our modelling technique focuses on massive halos, which are overdensity features. The prediction on local maxima with
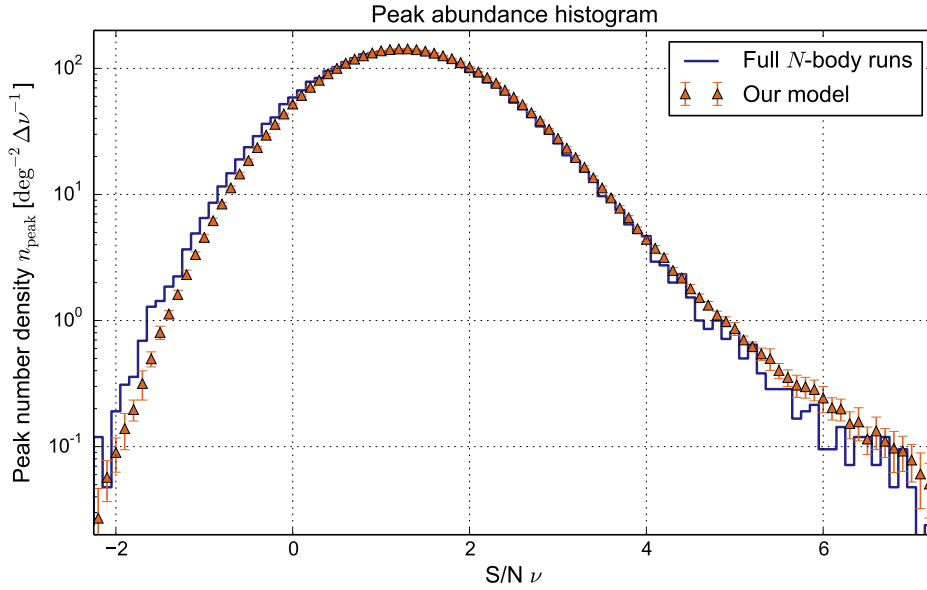
FIGURE 6.2: Similar to Fig. 5.9. Only N-body runs and our model are retained. The binwidth has been changed and the x range has been extended to -2. We can see that the number of negative peaks is systematically underestimated.

negative S/N has been found out to be incorrect, as shown by Fig. 6.2, which will further impact on values of $\boldsymbol{x}^{\mathrm{pct}}$. As a result, with real data, using directly $\boldsymbol{x}^{\mathrm{pct}}$ might introduce substantial bias. For this reason, $\boldsymbol{x}^{\mathrm{pct}}$ is included only for giving an idea about how much information we can extract from a data vector derived from the full inverse CDF, and $\boldsymbol{x}^{\mathrm{cut}}$ is a more realistic choice for future studies.

In Sect. 5.2.1, it is mentioned that although low-ν peaks might have better constraining power, it is not recommended to use for our model due to the lower mass cutoff M_{min} during mass sampling. For the same reason, I focus here on peaks with $\nu \geq 3$ and define $\boldsymbol{x}^{\mathrm{abd}}$, $\boldsymbol{x}^{\mathrm{pct}}$, and $\boldsymbol{x}^{\mathrm{cut}}$ as indicated in Table 6.2. Starting from choosing logarithmic separated percentiles for $\boldsymbol{x}^{\mathrm{cut}}$ (with $\nu_{\mathrm{min}} = 3$), I determine the average of the corresponding x_i over the 1000 realizations under $\boldsymbol{\pi}^{\mathrm{in}}$. Then, with these x_i, I define u_i for $\boldsymbol{x}^{\mathrm{pct}}$ and bins for $\boldsymbol{x}^{\mathrm{abd}}$ such that all three vectors represent approximately the same information. Of course, this coherence is only connected via the CDF under $\boldsymbol{\pi}^{\mathrm{in}}$ and might break down under another cosmology, but for the current studying framework, this choice can be considered reasonable.

Table 6.3 shows the correlation matrix for each of the data vectors under $\boldsymbol{\pi}^{\mathrm{in}}$. It reveals that the inter-component correlation for $\boldsymbol{x}^{\mathrm{abd}}$ is quite weak, and strong for $\boldsymbol{x}^{\mathrm{pct}}$ and $\boldsymbol{x}^{\mathrm{cut}}$. This suggests that the covariance should be included in likelihood analyses. The highest absolute value of off-diagonal terms does not exceed 17% in the case of $\boldsymbol{x}^{\mathrm{abd}}$. A similar result can be reproduced when binning peaks differently for $\boldsymbol{x}^{\mathrm{abd}}$.

6.1.4   How to qualify a constraint?

Performing comparison between estimators requires quantitative criteria on constraint contours. In this respect, two indicators have been found in literature. Inspired by Jain & Seljak

Label	abd				
Bins on ν	[3.0, 3.8[	[3.8, 4.5[	[4.5, 5.3[	[5.3, 6.2[	[6.2, $+\infty$[
x_i for $\boldsymbol{\pi}^{\mathrm{in}}$	330	91	39	18	15
Label	pct				
$\nu_{\min}$	$-\infty$				
u_i	0.969	0.986	0.994	0.997	0.999
x_i for $\boldsymbol{\pi}^{\mathrm{in}}$	3.5	4.1	4.9	5.7	7.0
Label	cut				
$\nu_{\min}$	3				
u_i	0.5	0.776	0.9	0.955	0.98
x_i for $\boldsymbol{\pi}^{\mathrm{in}}$	3.5	4.1	4.9	5.7	6.7

TABLE 6.2: Definition of $\boldsymbol{x}^{\mathrm{abd}}$, $\boldsymbol{x}^{\mathrm{pct}}$, and $\boldsymbol{x}^{\mathrm{cut}}$. The x_i corresponds to the number of peaks for the first case and S/N values for two others. Parameters such as $\nu_{\min}$ and u_i are used in Eq. (6.3). As an indication, their values for the input cosmology $\boldsymbol{\pi}^{\mathrm{in}}$ are also given. They were calculated by averaging over 1000 realizations.

(1997) and Maoli et al. (2001), the first quantity is the error on

$$\Sigma_8 \equiv \sigma_8 \left(\frac{\Omega_{\mathrm{m}}}{\Omega_0}\right)^\alpha. \tag{6.4}$$

In some papers, the pivot value Ω_0 is set to 0.3. Here, the convention of $\Omega_0 = 0.27$, adopted by Liu J. et al. (2015a) and Liu X. et al. (2015b), is used. This error, noted as $\Delta\Sigma_8$, is the "thickness" of the banana-shaped contour on the Ω_{m}-σ_8 plane tilted by the slope α. The value for the slope is usually taken from the best-fit result with the linear relation $\log \Sigma_8 = \log \sigma_8 + \alpha \log(\Omega_{\mathrm{m}}/0.27)$. Then, assuming this α value, one can transform the estimator into a function of Ω_{m} and Σ_8, marginalize over Ω_{m}, and derive $\Delta\Sigma_8$.

In this work, both frequentist and Bayesian analyses are performed. The frequentist approach proceeds what we call *likelihood-ratio test* (see e.g. Theorem 10.3.3 from Casella & Berger 2002). In a general mathematical context, this defines the *i-σ confidence region* as

$$\text{set of all } \boldsymbol{\pi} \text{ such that } 1 - p_i \geq \int_0^{-2\ln(\mathcal{L}(\boldsymbol{\pi})/\mathcal{L}_{\max})} \chi_f^2(t)\mathrm{d}t, \tag{6.5}$$

where χ_f^2 is the chi-squared distribution with f degrees of freedom (f is the dimension of $\boldsymbol{\pi}$) and p_i is the p-value corresponding to the i-σ significance, defined as

$$1 - p_i \equiv \frac{1}{\sqrt{2\pi}} \int_{-i}^{+i} \exp(-t^2/2)\mathrm{d}t. \tag{6.6}$$

For example, $1 - p_1 \approx 68.3\%$ and $1 - p_2 \approx 95.4\%$. The interpretation for the confidence region is that *the expected result from the considered model for any parameter inside the region should be at least better than the current observed data at the i-σ level*, where better refers to more probable. Here, f is just one since α has been fixed and the only parameter is Σ_8. The quantity $\Delta\Sigma_8$ is defined as the width of the 1-σ interval. In the case of $f = 1$, this is exactly the interval of Σ_8 such that $-2\ln(\mathcal{L}(\Sigma_8)/\mathcal{L}_{\max}) \leq 1$. The left panel of Fig. 6.3 shows an example of the likelihood-ratio test and the derived confidence interval on Σ_8.

$$
\texttt{abd} \quad
\begin{pmatrix}
1 & -0.05 & -0.09 & -0.08 & -0.16 \\
-0.05 & 1 & -0.05 & -0.01 & -0.12 \\
-0.09 & -0.05 & 1 & -0.04 & -0.11 \\
-0.08 & -0.01 & -0.04 & 1 & -0.06 \\
-0.16 & -0.12 & -0.11 & -0.06 & 1
\end{pmatrix}
$$

$$
\texttt{pct} \quad
\begin{pmatrix}
1 & 0.62 & 0.29 & 0.15 & 0.11 \\
0.62 & 1 & 0.58 & 0.36 & 0.25 \\
0.29 & 0.58 & 1 & 0.66 & 0.43 \\
0.15 & 0.36 & 0.66 & 1 & 0.59 \\
0.11 & 0.25 & 0.43 & 0.59 & 1
\end{pmatrix}
$$

$$
\texttt{cut} \quad
\begin{pmatrix}
1 & 0.58 & 0.31 & 0.20 & 0.15 \\
0.58 & 1 & 0.61 & 0.39 & 0.28 \\
0.31 & 0.61 & 1 & 0.65 & 0.47 \\
0.20 & 0.39 & 0.65 & 1 & 0.70 \\
0.15 & 0.28 & 0.47 & 0.70 & 1
\end{pmatrix}
$$

TABLE 6.3: Correlation matrices of $\boldsymbol{x}^{\mathrm{abd}}$, $\boldsymbol{x}^{\mathrm{pct}}$, and $\boldsymbol{x}^{\mathrm{cut}}$ in the input cosmology. For $\boldsymbol{x}^{\mathrm{abd}}$, the peak abundance is weakly correlated between bins.

For Bayesian inference, one seeks for the most probable region based on the likelihood function. This is called i-σ *credible region* and is defined as

$$
\text{set of all } \ \boldsymbol{\pi} \ \text{ such that } \ 1 - p_i \geq \int \mathrm{d}^f \boldsymbol{\pi}' \ \mathcal{L}(\boldsymbol{\pi}') \Theta(\mathcal{L}(\boldsymbol{\pi}') - \mathcal{L}(\boldsymbol{\pi})). \tag{6.7}
$$

As a result, $\Delta\Sigma_8$ is just the width of the 1-σ credible interval, as shown on the right panel of Fig. 6.3. I would like to highlight that the term *confidence* (region, contour, interval, etc.) always refers to a frequentist analysis, while the *ad hoc* word for a Bayesian approach is *credible*.

The second quantitative indicator for constraint contours is the figure of merit (FoM) proposed by Dietrich & Hartlap (2010). Analogous to the one from Albrecht et al. (2006), this is the inverse of the area of the 2-σ region on the Ω_{m}-σ_8 plane. Depending on the scenario, it can be either a confidence region (Eq. (6.5)) or a credible region (Eq. (6.7)). By defining $\Delta\Sigma_8$ and FoM, we can now analyze quantitatively the difference between estimators.

6.2 Gaussian likelihood and cosmology-dependent covariance

6.2.1 Formalisms

In this section, I show the parameter constraint contours from our model, and I further examine the so-called *cosmology-dependent-covariance effect* (CDC effect) on the Gaussian likelihood.

Following Sect. 6.1.1, the shape of the likelihood function is usually assumed to be a
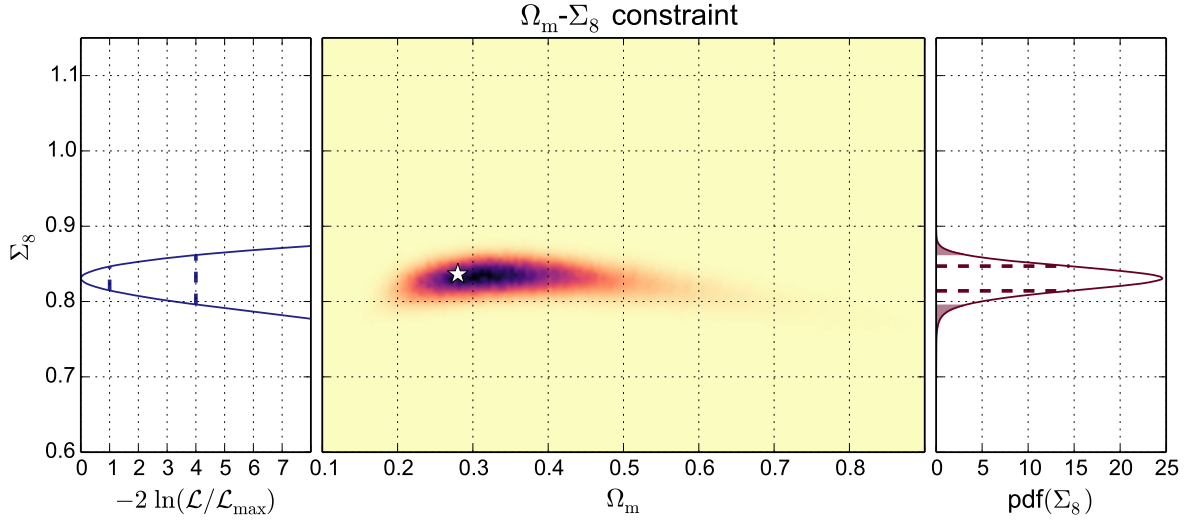
FIGURE 6.3: Middle panel: the likelihood value using $\boldsymbol{x}^{\mathrm{abd}}$ on the Ω_{m}-Σ_8 plane. The white star represents the input cosmology $\boldsymbol{\pi}^{\mathrm{in}}$. Since $\log\sigma_8$ and $\log\Omega_{\mathrm{m}}$ form an approximately linear degeneracy, the quantity $\Sigma_8 \equiv \sigma_8(\Omega_{\mathrm{m}}/0.27)^\alpha$ allows us to characterize the thickness. Right panel: the marginalized PDF of Σ_8. The dashed lines give the 1-σ interval (68.3%), while the borders of the shaded areas represent 2-σ limits (95.4%). Left panel: the log-value of the marginalized likelihood ratio. Dashed lines in the left panel give the corresponding value for 1- and 2-σ significance levels, respectively.

multivariate Gaussian, that is

$$\mathcal{L}_{\mathrm{g}}(\boldsymbol{\pi}) = \frac{1}{\sqrt{(2\pi)^d |\det \boldsymbol{C}(\boldsymbol{\pi})|}} \exp\left[-\frac{1}{2}\Delta\boldsymbol{x}^T(\boldsymbol{\pi})\cdot\boldsymbol{C}^{-1}(\boldsymbol{\pi})\cdot\Delta\boldsymbol{x}(\boldsymbol{\pi})\right], \tag{6.8}$$

where the label g stands for Gaussian, d denotes the dimension of $\boldsymbol{x}$, $\Delta\boldsymbol{x}(\boldsymbol{\pi}) \equiv \boldsymbol{x}^{\mathrm{obs}}-\boldsymbol{x}^{\mathrm{mod}}(\boldsymbol{\pi})$, the difference between the observation and the model prediction $\boldsymbol{x}^{\mathrm{mod}}(\boldsymbol{\pi})$, and $\boldsymbol{C}(\boldsymbol{\pi})$ is the covariance matrix for $\boldsymbol{x}^{\mathrm{mod}}(\boldsymbol{\pi})$.

In general, estimating $\boldsymbol{C}(\boldsymbol{\pi})$ could be difficult or expensive. As I have mentioned in Chap. 5, most studies call a relatively large amount of N-body simulations or resampling techniques such as bootstrap and jackknife to estimate the covariance. In order to make the estimation tractable, a common approximation is to assume that the covariance is independent from the cosmological parameters, so that $\boldsymbol{C}(\boldsymbol{\pi}) = \boldsymbol{C}(\boldsymbol{\pi}^{\mathrm{in}})$ is constant and the log-likelihood, defined as $L \equiv -2\ln\mathcal{L}$, can be written as

$$L_{\mathrm{cg}}(\boldsymbol{\pi}) = \mathrm{cst} + \Delta\boldsymbol{x}^T(\boldsymbol{\pi})\cdot\widehat{\boldsymbol{C}^{-1}}(\boldsymbol{\pi}^{\mathrm{in}})\cdot\Delta\boldsymbol{x}(\boldsymbol{\pi}), \tag{6.9}$$

where $\widehat{\boldsymbol{C}^{-1}}(\boldsymbol{\pi}^{\mathrm{in}})$ is the inverse covariance estimated at a chosen parameter $\boldsymbol{\pi}^{\mathrm{in}}$, and the constant comes from the determinant in Eq. (6.8) which is not important since we are always interested in the likelihood ratio. The remaining $\boldsymbol{\pi}$-dependent term in Eq. (6.9) is called "chi-squared term" (χ^2 term). The label cg is an abbreviation of constant-covariance Gaussian. If the covariance is estimated with resampling techniques, one would rather replace $\boldsymbol{C}(\boldsymbol{\pi}^{\mathrm{in}})$ by $\boldsymbol{C}(\boldsymbol{\pi}^{\mathrm{obs}})$, but the idea remains the same, which is treating it as an external parameter.

Does this constant-covariance hypothesis introduce an impact on parameter constraints? To examine this, two other estimators are proposed. The first one still neglects the determinant term in Eq. (6.8), but considers the χ^2 term parameter-dependent. In other words, this
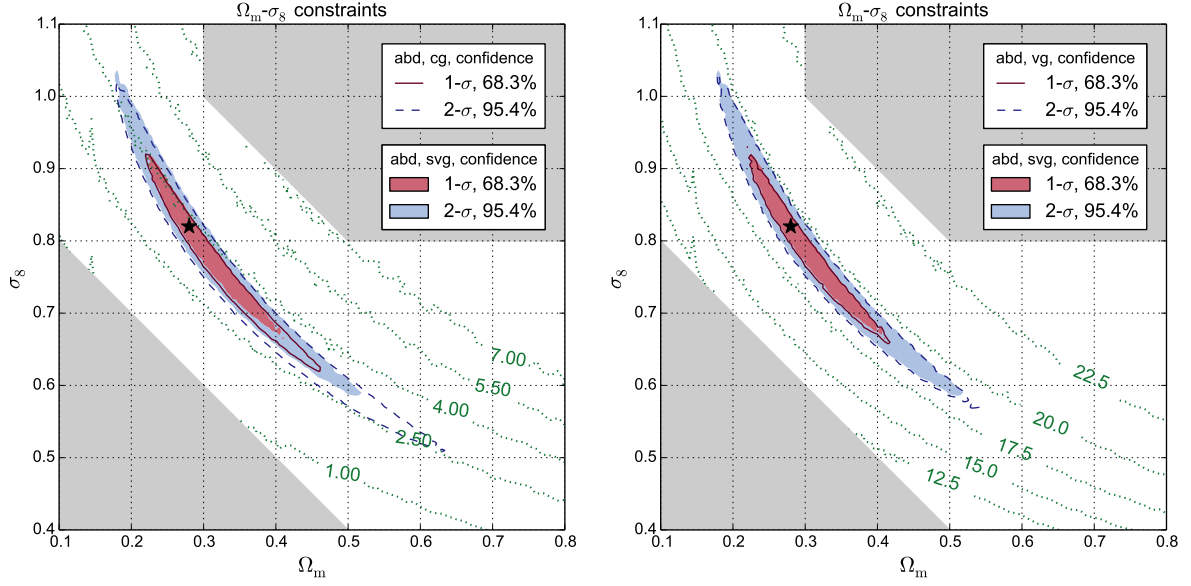
FIGURE 6.4: Confidence regions derived from L_{cg}, L_{svg}, and L_{vg} with $\boldsymbol{x}^{abd}$. The solid and dashed contours represent L_{cg} in the left panel and L_{vg} in the right panel, while the colored areas are from L_{svg}. The black star stands for $\boldsymbol{\pi}^{in}$ and gray areas represent the non-explored parameter space. The green dotted lines are different isolines, the variance $\widehat{C}_{55}$ of the bin with highest S/N in the left panel and $\ln(\det\widehat{C})$ for the right panel. The contour area is reduced by 22% when taking the CDC effect into account. The parameter-dependent determinant term does not contribute significantly.

semi-varying-covariance Gaussian log-likelihood (labelled svg) is

$$L_{svg}(\boldsymbol{\pi}) = \mathrm{cst} + \Delta\boldsymbol{x}^T(\boldsymbol{\pi}) \cdot \widehat{C^{-1}}(\boldsymbol{\pi}) \cdot \Delta\boldsymbol{x}(\boldsymbol{\pi}). \tag{6.10}$$

The second log-likelihood is a fully varying-covariance and Gaussian (labelled vg) function, mathematically given by

$$L_{vg}(\boldsymbol{\pi}) = \mathrm{cst} + \ln\left|\det\widehat{C}(\boldsymbol{\pi})\right| + \Delta\boldsymbol{x}^T(\boldsymbol{\pi}) \cdot \widehat{C^{-1}}(\boldsymbol{\pi}) \cdot \Delta\boldsymbol{x}(\boldsymbol{\pi}). \tag{6.11}$$

The likelihoods above are evaluated using the $N = 1000$ model realization set described in Sect. 6.1.2. The model prediction is the mean over N samples for each $\boldsymbol{\pi}$. Denote $\boldsymbol{x}^{(k)} = \left(x_1^{(k)}, \ldots, x_d^{(k)}\right)$ as the data vector of k-th realization, the unbiased estimators of the mean, the covariance matrix, and the inverse matrix (Hartlap et al. 2007) can be written respectively as

$$x_i^{mod} = \frac{1}{N}\sum_{k=1}^{N} x_i^{(k)}, \tag{6.12}$$

$$\widehat{C}_{ij} = \frac{1}{N-1}\sum_{k=1}^{N}\left(x_i^{(k)} - x_i^{mod}\right)\left(x_j^{(k)} - x_j^{mod}\right), \tag{6.13}$$

$$\widehat{C^{-1}} = \frac{N-d-2}{N-1}\,\widehat{C}^{-1}. \tag{6.14}$$

6.2.2 The chi-squared term

What is the impact of the CDC effect on the χ^2 term? The left panel of Fig. 6.4 shows the comparison between confidence regions derived from L_{cg} and L_{svg}. The data vector is $\boldsymbol{x}^{abd}$.
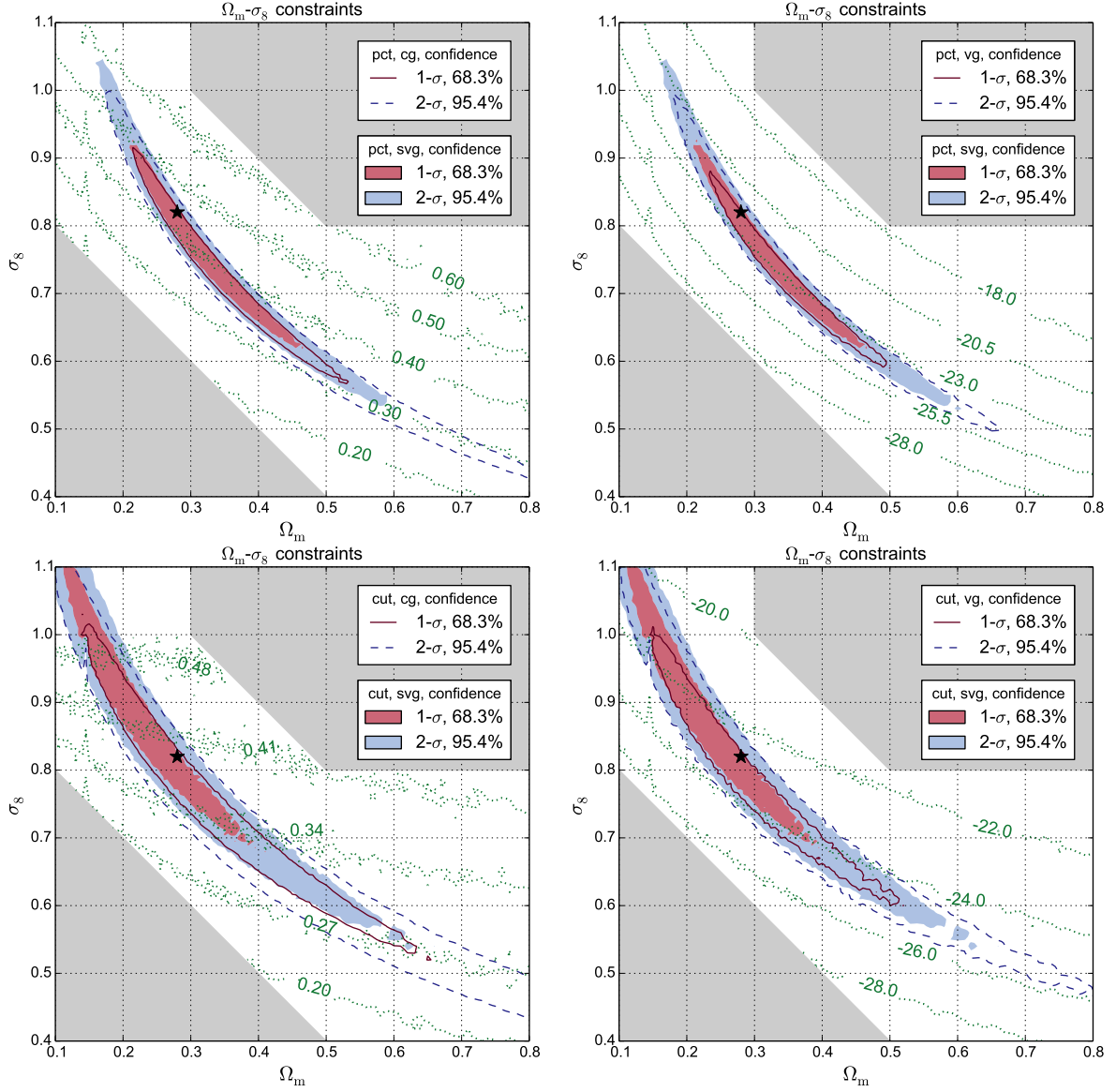
FIGURE 6.5: Similar to Fig. 6.4. Confidence regions with $\boldsymbol{x}^{\mathrm{pct}}$ and $\boldsymbol{x}^{\mathrm{cut}}$. Both upper panels are drawn with $\boldsymbol{x}^{\mathrm{pct}}$ and both lower panels with $\boldsymbol{x}^{\mathrm{cut}}$. Both left panels are the comparison between L_{cg} and L_{svg}, and both right panels between L_{svg} and L_{vg}.

One can observe that the constraint contours have been shifted. As shown in Table 6.3, the off-diagonal correlation coefficients are weak. Therefore, the variation in diagonal terms of $\widehat{\boldsymbol{C}}$ plays a major role in the size of confidence regions. In the same figure, the isolines for $\widehat{C}_{55}$ are also drawn. These isolines cross the Ω_{m}-σ_8 degeneracy lines from L_{cg} and thus shrink the confidence region. More detailed examinations reveal that the isolines for $\widehat{C}_{11}$ and $\widehat{C}_{22}$ are noisy and that those for $\widehat{C}_{33}$ and $\widehat{C}_{44}$ coincide well with the original degeneracy direction.

Table 6.4 shows the values of both indicators for different likelihoods. One can observe that using L_{svg} significantly improves the constraints by 24% in terms of FoM compared to L_{cg}. Regarding $\Delta\Sigma_8$, the improvement is weak. As a result, using varying covariance matrices breaks down the degeneracy partially and shrinks the contour length, but does not effectively reduce the thickness.

	$\boldsymbol{x}^{\mathrm{abd}}$		$\boldsymbol{x}^{\mathrm{pct}}$		$\boldsymbol{x}^{\mathrm{cut}}$	
	$\Delta\Sigma_8$	FoM	$\Delta\Sigma_8$	FoM	$\Delta\Sigma_8$	FoM
L_{cg}	0.032	46	0.037	31	0.065	13
L_{svg}	0.031	57	0.032	42	0.054	21
L_{vg}	0.031	56	0.032	43	0.052	18
L_{cc}	0.032	43	0.038	33	0.056	13
L_{vc}	0.033	52	0.034	39	0.058	16
L_{true}	0.033	54	0.035	39	0.058	17
p-value	0.035	39	0.037	27	0.067	12

TABLE 6.4: Values of $\Delta\Sigma_8$ and the figure of merit (FoM) for confidence regions from different analysis approaches. L_{cg}, L_{svg}, and L_{vg} are introduced in Sect. 6.2.1, L_{cc} and L_{vc} in Sect. 6.3.2, and L_{true} and p-value in Sect. 6.4. In each case, $\boldsymbol{x}^{\mathrm{abd}}$, $\boldsymbol{x}^{\mathrm{pct}}$, or $\boldsymbol{x}^{\mathrm{cut}}$ are taken respectively as data vector as indicated in the table rows.

	$\boldsymbol{x}^{\mathrm{abd}}$		$\boldsymbol{x}^{\mathrm{pct}}$		$\boldsymbol{x}^{\mathrm{cut}}$	
	$\Sigma_8\ {}^{+1\sigma}_{-1\sigma}$	α	$\Sigma_8\ {}^{+1\sigma}_{-1\sigma}$	α	$\Sigma_8\ {}^{+1\sigma}_{-1\sigma}$	α
L_{cg}	$0.831^{+0.016}_{-0.016}$	0.54	$0.822^{+0.018}_{-0.019}$	0.54	$0.800^{+0.030}_{-0.035}$	0.45
L_{svg}	$0.831^{+0.016}_{-0.015}$	0.52	$0.820^{+0.015}_{-0.015}$	0.51	$0.800^{+0.031}_{-0.023}$	0.40
L_{vg}	$0.829^{+0.015}_{-0.015}$	0.52	$0.819^{+0.015}_{-0.016}$	0.52	$0.800^{+0.024}_{-0.028}$	0.42
L_{cc}	$0.830^{+0.016}_{-0.016}$	0.54	$0.825^{+0.018}_{-0.020}$	0.54	$0.807^{+0.025}_{-0.031}$	0.46
L_{vc}	$0.829^{+0.016}_{-0.016}$	0.52	$0.823^{+0.016}_{-0.019}$	0.53	$0.798^{+0.029}_{-0.029}$	0.44
L_{true}	$0.828^{+0.018}_{-0.015}$	0.53	$0.823^{+0.015}_{-0.020}$	0.53	$0.800^{+0.028}_{-0.030}$	0.44
p-value	$0.835^{+0.016}_{-0.019}$	0.54	$0.823^{+0.018}_{-0.018}$	0.54	$0.798^{+0.032}_{-0.034}$	0.45

TABLE 6.5: Best fits of (Σ_8, α) for confidence regions from different analysis approaches. L_{cg}, L_{svg}, and L_{vg} are introduced in Sect. 6.2.1, L_{cc} and L_{vc} in Sect. 6.3.2, and L_{true} and p-value in Sect. 6.4. Readers should note that the values for Σ_8 are indicative since no real observational data are used.

The same constraints derived from two other data vectors $\boldsymbol{x}^{\mathrm{pct}}$ and $\boldsymbol{x}^{\mathrm{cut}}$ can be found in the left panel of Fig. 6.5. A similar CDC effect exists for both. One can remark that $\boldsymbol{x}^{\mathrm{pct}}$ has less constraining power than $\boldsymbol{x}^{\mathrm{abd}}$. Also, $\boldsymbol{x}^{\mathrm{cut}}$ is outperformed by both other data vectors. This difference is due to the cutoff value ν_{min}. Introducing a cutoff at $\nu_{\mathrm{min}} = 3$ decreases the total number of peaks and amplifies the fluctuation of high-peak values in the CDF. When we use percentiles to define the data vector, the distribution of each component of $\boldsymbol{x}^{\mathrm{cut}}$ becomes wider than the one of the corresponding component of $\boldsymbol{x}^{\mathrm{pct}}$, and this greater scatter in the CDF enlarges the contours. However, the cutoff also introduces a tilt for the contours. Table 6.5 shows the best fit alpha for the different cases. The difference in the tilt could be a useful tool for improving the constraining power, as demonstrated by Dietrich & Hartlap (2010). Nevertheless, I do not explore any joint analysis since Dietrich & Hartlap (2010) consider peak counts and power spectrum while here, $\boldsymbol{x}^{\mathrm{abd}}$ and $\boldsymbol{x}^{\mathrm{cut}}$ contain basically the same information.

	$\boldsymbol{x}^{\mathrm{abd}}$		$\boldsymbol{x}^{\mathrm{pct}}$		$\boldsymbol{x}^{\mathrm{cut}}$	
	$\Delta\Sigma_8$	FoM	$\Delta\Sigma_8$	FoM	$\Delta\Sigma_8$	FoM
L_{cg}	0.033	43	0.038	31	0.066	15
L_{svg}	0.031	53	0.033	41	0.056	20
L_{vg}	0.031	53	0.032	40	0.055	18
L_{cc}	0.033	40	0.040	30	0.071	14
L_{vc}	0.033	47	0.035	36	0.060	16
L_{true}	0.034	49	0.036	36	0.061	17
ABC	0.056	31	0.044	33	0.068	16

TABLE 6.6: Similar to Table 6.4, but for credible regions. The row ABC represents the results from Sect. 7.2.4.

	$\boldsymbol{x}^{\mathrm{abd}}$		$\boldsymbol{x}^{\mathrm{pct}}$		$\boldsymbol{x}^{\mathrm{cut}}$	
	$\Sigma_8 \, {}^{+1\sigma}_{-1\sigma}$	α	$\Sigma_8 \, {}^{+1\sigma}_{-1\sigma}$	α	$\Sigma_8 \, {}^{+1\sigma}_{-1\sigma}$	α
L_{cg}	$0.831^{+0.017}_{-0.016}$	0.54	$0.822^{+0.018}_{-0.020}$	0.54	$0.800^{+0.030}_{-0.036}$	0.45
L_{svg}	$0.831^{+0.016}_{-0.015}$	0.52	$0.820^{+0.016}_{-0.017}$	0.51	$0.800^{+0.032}_{-0.024}$	0.40
L_{vg}	$0.829^{+0.015}_{-0.015}$	0.52	$0.819^{+0.015}_{-0.017}$	0.52	$0.800^{+0.025}_{-0.029}$	0.42
L_{cc}	$0.830^{+0.017}_{-0.017}$	0.54	$0.825^{+0.018}_{-0.022}$	0.54	$0.807^{+0.030}_{-0.041}$	0.46
L_{vc}	$0.829^{+0.016}_{-0.016}$	0.52	$0.823^{+0.016}_{-0.019}$	0.53	$0.798^{+0.030}_{-0.030}$	0.44
L_{true}	$0.828^{+0.019}_{-0.015}$	0.53	$0.823^{+0.015}_{-0.021}$	0.53	$0.800^{+0.030}_{-0.032}$	0.44
ABC	$0.819^{+0.030}_{-0.025}$	0.50	$0.817^{+0.022}_{-0.022}$	0.51	$0.799^{+0.034}_{-0.034}$	0.42

TABLE 6.7: Similar to Table 6.5, but for credible regions. The row ABC represents the results from Sect. 7.2.4. The values for Σ_8 are only indicative.

6.2.3 The determinant term

The right panel of Fig. 6.4 shows the comparison between L_{svg} and L_{vg} with the data vector $\boldsymbol{x}^{\mathrm{abd}}$. The difference between contours is minor. This shows that adding the determinant term does not result in significant changes of the parameter constraints. The isolines from $\ln(\det \widehat{\boldsymbol{C}})$ illustrate well this phenomenon. The gradients, this time, are perpendicular to the degeneracy lines. Meanwhile, including the determinant makes the contours slightly larger, but almost negligibly so. Quantitatively, the total improvement in the contour area compared to L_{cg} is 22%.

However, a different change is seen for $\boldsymbol{x}^{\mathrm{pct}}$ and $\boldsymbol{x}^{\mathrm{cut}}$ (Fig. 6.5). Adding the determinant to the likelihood computed from these observables induces a shift of contours toward the higher Ω_{m} area. In the case of $\boldsymbol{x}^{\mathrm{cut}}$, this shift compensates for the contour offset from the varying χ^2 term, but does not improve either $\Delta\Sigma_8$ or FoM significantly, as shown in Table 6.4. As a result, using the Gaussian likelihood, the total CDC effect can be summed up as an improvement of at least 14% in terms of thickness and 38% in terms of area.

Overall, the results from Bayesian inference are very similar to the likelihood-ratio test. Thus, only $\Delta\Sigma_8$ and FoM are shown in Table 6.6 and best fits in Table 6.7. I would like to

recall that a similar analysis has been done by Eifler et al. (2009) on shear covariances. This work agrees with their conclusions: a relatively large impact from the χ^2 term and negligible change from the determinant term. However, the total CDC effect is more significant in the peak-count framework than for the power spectrum.

6.3 Copula analysis

6.3.1 What is copula?

Consider a multivariate joint distribution $P(x_1, \ldots, x_d)$. In general, P could be far from Gaussian so that imposing the Gaussian likelihood could induce biases. The idea of the copula technique is to evaluate the likelihood in a new data space in which the Gaussian approximation is better. Using a change of variables, individual marginalized distributions of P can be approximated to Gaussian ones. This is achieved by a series of successive 1D, axis-wise transformations. In general, the multivariate Gaussianity of the transformed distribution is still **not** guaranteed. However, in some cases, this transformation tunes the distribution and makes it "more Gaussian", so that evaluating the likelihood in the tuned space is more realistic (Benabed et al. 2009; Sato et al. 2011).

From Sklar's theorem (Sklar 1959), any multivariate distribution $P(x_1, \ldots, x_d)$ can be decomposed into the copula density multiplied by marginalized distributions. A comprehensible and elegant demonstration is given by Rüschendorf (2009). Readers are also encouraged to follow Scherrer et al. (2010) for detailed physical interpretations and Sato et al. (2011) for a very pedagogical derivation of the Gaussian copula transform.

Let $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $P(\boldsymbol{x})$ a multivariate distribution. For all i, the marginalized distribution is naturally

$$P_i(x_i) \equiv \int P(\boldsymbol{x}) \prod_{j \neq i} \mathrm{d}x_j, \tag{6.15}$$

and its CDF leads

$$F_i(x_i) \equiv \int_{-\infty}^{x_i} P_i(x') \mathrm{d}x'. \tag{6.16}$$

We can also define the d-dimensional cumulative distribution of P as

$$F(\boldsymbol{x}) \equiv \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} P(x'_1, \ldots, x'_d) \mathrm{d}x'_1 \cdots \mathrm{d}x'_d. \tag{6.17}$$

Now, Sklar's theorem indicates that there exists an unique d-dimensional distribution function C defined on $[0,1]^d$ with uniform marginals such that

$$F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{6.18}$$

This function C is called the *copula*. In other words, by defining $u_i \equiv F_i(x_i)$ and $x_i = F_i^{-1}(u_i)$, the copula is defined as

$$C(\boldsymbol{u}) \equiv F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)). \tag{6.19}$$

We can compute the *copula density*, $c(\boldsymbol{u})$, given by

$$c(\boldsymbol{u}) \equiv \frac{\partial C(\boldsymbol{u})}{\partial u_1 \cdots \partial u_d} = \frac{\partial F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{\partial x_1 \cdots \partial x_d} \cdot \frac{\partial x_1}{\partial u_1} \cdots \frac{\partial x_d}{\partial u_d}, \tag{6.20}$$

and this results in

$$P(\boldsymbol{x}) = c(\boldsymbol{u})P_1(x_1)\cdots P_d(x_d), \tag{6.21}$$

which means that *any multivariate distribution can be decomposed into the product of the copula density and marginalized distributions.*

On the other hand, let $q_i \equiv \Phi_i^{-1}(u_i)$, where Φ_i is the CDF of the normal distribution with the same mean μ_i and variance σ_i^2 as the law P_i. Mathematically, this leads

$$\Phi_i(q_i) \equiv \int_{-\infty}^{q_i} \phi_i(q')\mathrm{d}q', \tag{6.22}$$

$$\phi_i(q_i) \equiv \frac{1}{\sqrt{2\pi\sigma_i^2}} \, \exp\left[-\frac{(q_i - \mu_i)^2}{2\sigma_i^2}\right]. \tag{6.23}$$

The relation between $\boldsymbol{q}$ and $\boldsymbol{x}$ is just a change of variables, so there exists a new joint PDF in the $\boldsymbol{q}$-space which corresponds to P in $\boldsymbol{x}$-space, say P'. Then, we have $P'(\boldsymbol{q}) = P(\boldsymbol{x})$. The marginal PDF and CDF of P' are nothing but ϕ_i and Φ_i, respectively. Thus, applying Eq. (6.21) to P' and ϕ_i, one gets

$$P'(\boldsymbol{q}) = c(\boldsymbol{u})\phi_1(q_1)\cdots\phi_d(q_d). \tag{6.24}$$

By uniqueness of c, the copula densities in Eqs. (6.21) and (6.24) turn out to be the same. Thus, we obtain

$$P(\boldsymbol{x}) = P'(\boldsymbol{q})\frac{P_1(x_1)\cdots P_d(x_d)}{\phi_1(q_1)\cdots\phi_d(q_d)}. \tag{6.25}$$

Readers should note that the marginals of P' are identical to the ones of a multivariate Gaussian distribution ϕ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{C}$, where $\boldsymbol{C}$ is the covariance matrix of $\boldsymbol{x}$. The PDF of this Gaussian distribution is

$$\phi(\boldsymbol{q}) \equiv \frac{1}{\sqrt{(2\pi)^d|\det\boldsymbol{C}|}} \exp\left[-\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}(q_i - \mu_i)C_{ij}^{-1}(q_j - \mu_j)\right]. \tag{6.26}$$

Finally, by approximating P' to ϕ, one gets the *Gaussian copula transform*:

$$P(\boldsymbol{x}) = \phi(\boldsymbol{q})\frac{P_1(x_1)\cdots P_d(x_d)}{\phi_1(q_1)\cdots\phi_d(q_d)}. \tag{6.27}$$

Why is it better to calculate the likelihood in this way? In the usual case, people often approximate the unknown shape of $P(\boldsymbol{x})$ to a normal distribution: $P(\boldsymbol{x}) \approx \phi(\boldsymbol{x})$. Here, by applying the Gaussian copula transform, we carry out this approximation into the new space of $\boldsymbol{q}$: $P'(\boldsymbol{q}) \approx \phi(\boldsymbol{q})$. Since $q_i = \Phi_i^{-1}(F_i(x_i))$, at least the marginals of $P'(\boldsymbol{q})$ are strictly Gaussian. And Eq. (6.27) gives the corresponding value in $\boldsymbol{x}$-space while taking the Gaussian approximation $P'(\boldsymbol{q}) \approx \phi(\boldsymbol{q})$ in $\boldsymbol{q}$-space. This is the reason that the copula transform makes the distribution "more Gaussian".

However, in some specific cases, the copula has no effect at all. Consider the following example:

$$f(x, y) = 2\phi(x, y)\Theta(xy), \quad \text{where} \quad \phi(x, y) = \frac{1}{\sqrt{(2\pi)^2}} \, \exp\left(-\frac{x^2 + y^2}{2}\right). \tag{6.28}$$
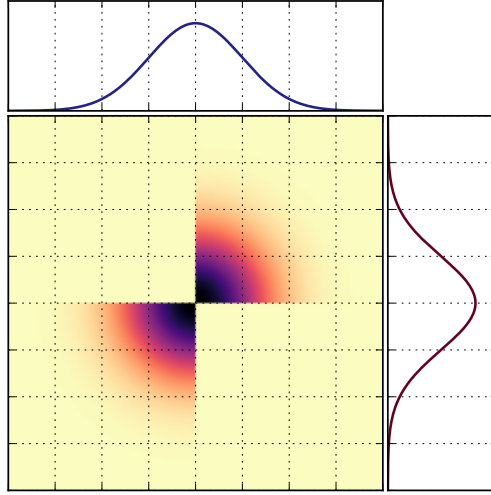
FIGURE 6.6: An example that the copula transform is insensitive. The multivariate distribution is not very "Gaussian", but its two marginals are both perfectly Gaussian, so the copula will not do adjustment.

Here, ϕ is just the 2D standard normal distribution and Θ is the Heaviside step function. The value of f is 2 times ϕ_2 if x and y have the same sign and 0 otherwise. The marginal PDF of f and ϕ_2 turn out to be the same (see Fig. 6.6). As a result, the Gaussian copula transform does nothing for f. The copula likelihood of f is exactly the original Gaussian likelihood, while f remains "highly" non Gaussian. This example proves that the copula is not the ultimate solution. However, if we do not have any prior knowledge, then the result with the copula transformation should be at least as good as the usual Gaussian likelihood.

6.3.2 Copula likelihood

We can now write down the copula likelihood. By applying Eq. (6.27) to $P(\boldsymbol{x}^{\mathrm{obs}}|\boldsymbol{\pi})$, x_i becomes x_i^{obs} and q_i becomes q_i^{obs} similarly. The *copula likelihood* $\mathcal{L}_{\mathrm{c}}$ is then

$$\mathcal{L}_{\mathrm{c}}(\boldsymbol{\pi}) = \frac{1}{\sqrt{(2\pi)^d |\det \boldsymbol{C}|}} \times \exp\left[-\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\left(q_i^{\mathrm{obs}} - x_i^{\mathrm{mod}}\right) C_{ij}^{-1}\left(q_j^{\mathrm{obs}} - x_j^{\mathrm{mod}}\right)\right]$$

$$\times \prod_{i=1}^{d}\left[\frac{1}{\sqrt{2\pi\sigma_i^2}}\,\exp\left(-\frac{(q_i^{\mathrm{obs}} - x_i^{\mathrm{mod}})^2}{2\sigma_i^2}\right)\right]^{-1} \times \prod_{i=1}^{d} P_i(x_i^{\mathrm{obs}}), \qquad (6.29)$$

where μ_i, the mean of q_i, is also the mean of x_i thus is replaced by x_i^{mod}. By detailing the dependency on $\boldsymbol{\pi}$ for all quantities and $\hat{\ }$ for estimated terms, the varying-covariance copula log-likelihood (labelled vc) is

$$L_{\mathrm{vc}}(\boldsymbol{\pi}) = \mathrm{cst} + \ln\left|\det \widehat{\boldsymbol{C}}(\boldsymbol{\pi})\right| + \sum_{i=1}^{d}\sum_{j=1}^{d}\left[q_i^{\mathrm{obs}}(\boldsymbol{\pi}) - x_i^{\mathrm{mod}}(\boldsymbol{\pi})\right]\widehat{C_{ij}^{-1}}(\boldsymbol{\pi})\left[q_j^{\mathrm{obs}}(\boldsymbol{\pi}) - x_j^{\mathrm{mod}}(\boldsymbol{\pi})\right]$$

$$- 2\sum_{i=1}^{d}\ln\hat{\sigma}_i(\boldsymbol{\pi}) - \sum_{i=1}^{d}\left(\frac{q_i^{\mathrm{obs}}(\boldsymbol{\pi}) - x_i^{\mathrm{mod}}(\boldsymbol{\pi})}{\hat{\sigma}_i(\boldsymbol{\pi})}\right)^2 - 2\sum_{i=1}^{d}\ln\hat{P}_i(x_i^{\mathrm{obs}}|\boldsymbol{\pi}). \qquad (6.30)$$

Here, $\hat{P}_i(x_i^{\mathrm{obs}}|\boldsymbol{\pi})$ should be understood as a one-point evaluation of $\hat{P}_i(\,\cdot\,|\boldsymbol{\pi})$ on x_i^{obs}, and $\hat{P}_i(\,\cdot\,|\boldsymbol{\pi})$ is the i-th marginal PDF in the $\boldsymbol{x}$ space, which is $\boldsymbol{\pi}$-dependent. This can be computed
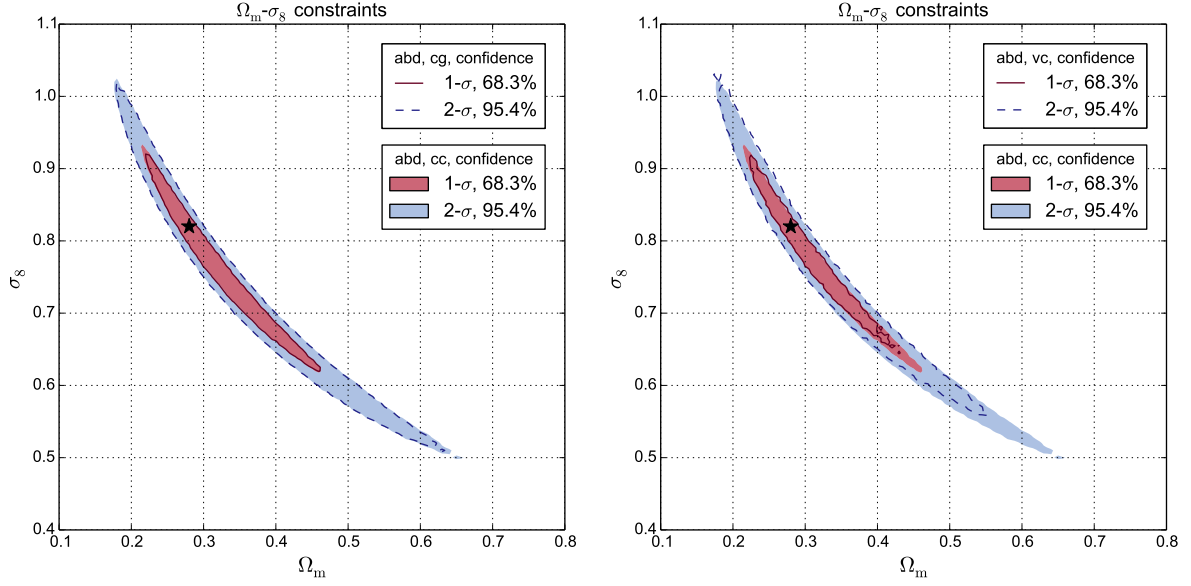
FIGURE 6.7: Confidence regions derived from copula analyses. Left panel: comparison between L_{cg} (solid and dashed contours) and L_{cc} (colored areas). Right panel: comparison between L_{cc} (colored areas) and L_{vc} (solid and dashed contours). The evolution tendency from L_{cc} to L_{vc} is similar to the evolution from L_{cg} to L_{vg}. The data vector is $\boldsymbol{x}^{\mathrm{abd}}$.

by applying kernel density estimation (KDE, see Sect. 6.4.1 and Eq. (6.33)) on the realization set. The quantities $x_i^{\mathrm{mod}}(\boldsymbol{\pi})$, $\hat{\sigma}_i(\boldsymbol{\pi})$, and $\widehat{C_{ij}^{-1}}(\boldsymbol{\pi})$ are estimated following Eqs. (6.12), (6.13), and (6.14). Finally, $q_i^{\mathrm{obs}}(\boldsymbol{\pi}) = \Phi_i^{-1}(\hat{F}_i(x_i^{\mathrm{obs}}|\boldsymbol{\pi}))$, where $\hat{F}_i(\,\cdot\,|\boldsymbol{\pi})$ is the CDF corresponding to $\hat{P}_i(\,\cdot\,|\boldsymbol{\pi})$ and Φ_i also depends implicitly on $\boldsymbol{\pi}$ via μ_i and $\hat{\sigma}_i$.

Similar to Sect. 6.2.1, we can also study the CDC effect on the copula likelihood. Obtained straightforwardly from Eq. (6.30), the constant-covariance copula log-likelihoods (labelled cc) is

$$
L_{\mathrm{cc}}(\boldsymbol{\pi}) = \mathrm{cst} + \sum_{i=1}^{d}\sum_{j=1}^{d}\left[q_i^{\mathrm{obs}}(\boldsymbol{\pi}) - x_i^{\mathrm{mod}}(\boldsymbol{\pi})\right]\widehat{C_{ij}^{-1}}\left[q_j^{\mathrm{obs}}(\boldsymbol{\pi}) - x_j^{\mathrm{mod}}(\boldsymbol{\pi})\right]
$$
$$
- \sum_{i=1}^{d}\left(\frac{q_i^{\mathrm{obs}}(\boldsymbol{\pi}) - x_i^{\mathrm{mod}}(\boldsymbol{\pi})}{\hat{\sigma}_i}\right)^2 - 2\sum_{i=1}^{d}\ln\hat{P}_i\left(x_i^{\mathrm{obs}} - x_i^{\mathrm{mod}}(\boldsymbol{\pi})\right), \qquad (6.31)
$$

where $\hat{P}_i(\,\cdot\,)$ denotes the zero-mean marginal PDF, assuming that the distribution of each x_i around its mean value is cosmology-independent. This PDF, $\hat{\sigma}_i$, and $\widehat{C_{ij}^{-1}}$ are estimated with the 1000 realizations under $\boldsymbol{\pi}^{\mathrm{in}}$.

6.3.3 Results

The comparisons between different likelihoods are shown in Fig. 6.7. The left panel focuses on the impact from the copula transformation using $\boldsymbol{x}^{\mathrm{abd}}$. It illustrates the confidence contour evolution from L_{cg} to L_{cc}. It turns out that the Gaussian likelihood is a very good approximation for lensing peak-counts. Presented in Table 6.4, quantitative results reveal that L_{cg} provides slightly more optimistic Ω_{m}-σ_8 constraints. I would like to underline that the effect of the copula transform is ambiguous, and both tighter or wider constraints are possible. This

has already shown by Sato et al. (2011), who found that the Gaussian likelihood underestimates the constraint power for low ℓ of the lensing power spectrum and overestimates it for high ℓ.

The right panel of Fig. 6.7 focuses on the CDC effect, comparing confidence regions between L_{cc} and L_{vc} using $\boldsymbol{x}^{\mathrm{abd}}$. When the CDC is taken into account for the copula transform, the parameter constraints are submitted compared to a similar change to the Gaussian likelihood. Tighter constraints are obtained from L_{vc} than from L_{cc}. The results from $\boldsymbol{x}^{\mathrm{pct}}$ and $\boldsymbol{x}^{\mathrm{cut}}$, as well as those on credible regions, lead to the same conclusion.

In summary, the copula likelihood with a fully varying covariance, L_{vc}, is closer to the truth than the usual Gaussian likelihood and is easy to compute. Taking this weaker approximation into account, the constraining power could be improved by at least 10% in terms of FoM compared to L_{cg}, the usual estimator.

6.4 Toward non-parametric estimation

6.4.1 True likelihood

This section presents two ways to obtain parameter constraints with even less restrictive assumptions. The first one is to use the "true" likelihood: if our model provides the PDF of $\boldsymbol{x}$, why not use it directly? The original definition of likelihood is $\mathcal{L}(\boldsymbol{\pi}) = P(\boldsymbol{x}^{\mathrm{obs}}|\boldsymbol{\pi})$, so we can write the true log-likelihood as

$$L_{\mathrm{true}}(\boldsymbol{\pi}) = -2\ln\hat{P}(\boldsymbol{x}^{\mathrm{obs}}|\boldsymbol{\pi}), \tag{6.32}$$

where $\hat{P}$ is an estimation of the PDF.

In this work, $\hat{P}$ is estimated with the realization set described in Sect. 6.1.2, using kernel density estimation (KDE). Noting $\boldsymbol{x}^{(k)}$ for $k = 1, \ldots, N$, the N samples of the data vector under $\boldsymbol{\pi}$, the multivariate density of $\boldsymbol{x}$ is

$$\hat{P}(\boldsymbol{x}|\boldsymbol{\pi}) = \frac{1}{N}\sum_{k=1}^{N} W_{\boldsymbol{H}}\left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right) \tag{6.33}$$

with

$$W_{\boldsymbol{H}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d|\det\boldsymbol{H}|}}\exp\left[-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}^{-1}\boldsymbol{x}\right] \tag{6.34}$$

and the bandwidth matrix $\boldsymbol{H}$ is given by Silverman's rule (Silverman 1986) as

$$\sqrt{H_{ij}} = \begin{cases} \left[\dfrac{4}{(d+2)N}\right]^{\frac{1}{d+4}}\hat{\sigma}_i & \text{if } i = j, \\[3mm] 0 & \text{otherwise,} \end{cases} \tag{6.35}$$

where d is the dimension of the data vector.

The left panel of Fig. 6.8 shows the confidence contours from L_{true} with $\boldsymbol{x}^{\mathrm{abd}}$. Readers can notice that this constraint is very noisy. This is due to a relatively low number of realizations which fails to stabilize the probability estimation and prevents us from making
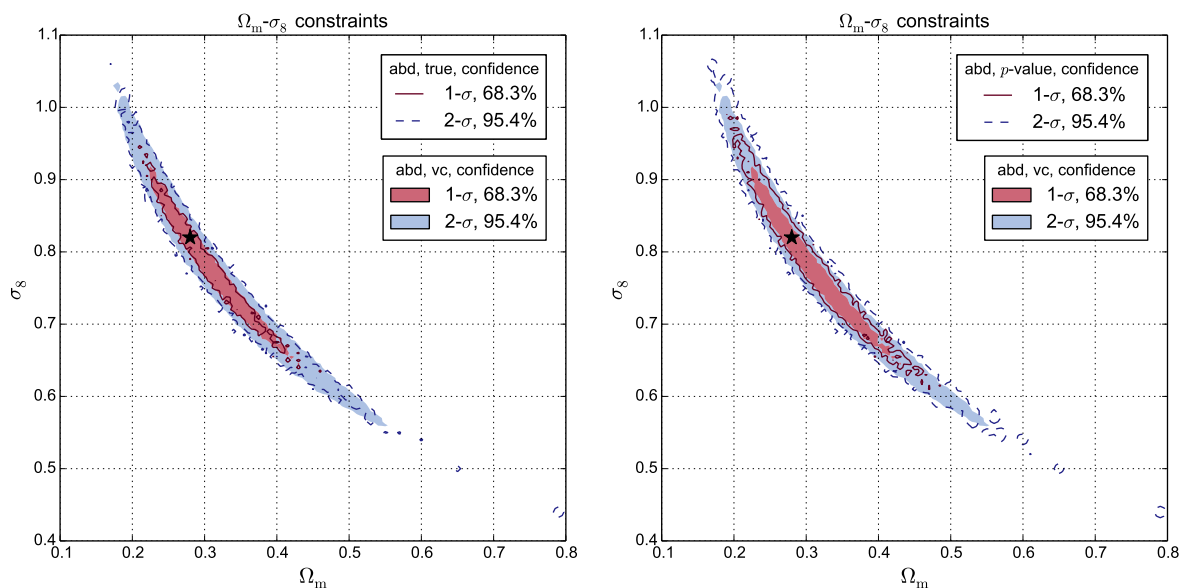
FIGURE 6.8: Left panel: confidence regions derived from L_{vc} (colored areas) and L_{true} (solid and dashed contours). Right panel: confidence regions derived from L_{vc} (colored areas) and p-value analysis (solid and dashed contours). The contours from L_{true} and p-value analysis are noisy due to a relatively low N. We can observe that L_{vc} and L_{true} yield very similar results. The data vector is $\boldsymbol{x}^{abd}$.

definite conclusions. However, when we put the constraint from L_{true} and the one from L_{vc} together, both results agree well and thus suggest that we may substitute L_{true} with L_{vc}, which bypasses the drawback of noisy estimation. Quantitative results from $\Delta\Sigma_8$ and FoM (Table 6.4) also suggest the same agreement. For constraints with $\boldsymbol{x}^{pct}$, $\boldsymbol{x}^{cut}$, and credible regions, I have found similar results and conclusions.

6.4.2 p-**value analysis**

Another non-parametric technique is the p-value analysis. Without passing by the likelihood, it gives directly the significance level of the observation for a model. For a observation $\boldsymbol{x}^{obs}$, the p-value associated to a parameter vector $\boldsymbol{\pi}$ is defined as

$$1 - p(\boldsymbol{\pi}) \equiv \int \mathrm{d}^d \boldsymbol{x} \ \hat{P}(\boldsymbol{x}|\boldsymbol{\pi}) \times \Theta\left(\hat{P}(\boldsymbol{x}|\boldsymbol{\pi}) - \hat{P}(\boldsymbol{x}^{obs}|\boldsymbol{\pi})\right), \tag{6.36}$$

where Θ denotes the Heaviside step function. The integral extends over the region where $\boldsymbol{x}$ is more probable than $\boldsymbol{x}^{obs}$, as shown by the left panel of Fig. 6.9. Thus, the proper interpretation of Eq. (6.36) is that if we generated N universes, then at least $(1 - p)N$ of them should have an observational result "better" than $\boldsymbol{x}^{obs}$, and "better" refers to more probable.

As in Sect. 6.4.1, KDE is performed to estimate the multivariate PDF and Eq. (6.36) is numerically integrated to obtain the p-value. Monte Carlo integration is used for evaluating the five-dimensional integrals in this case (consequence of the curse of dimensionality). To reach i-σ significance level, the p-value should reach at least p_i, which is given by Eq. (6.6). And as the right panel Fig. 6.9 shows, this provides a straightforward way to distinguish different cosmological models.

The result from p-value analysis with $\boldsymbol{x}^{abd}$ is presented in the right panel of Fig. 6.8 and Table 6.4. The contours seem to be larger than L_{true}, and are also quite noisy. This time,
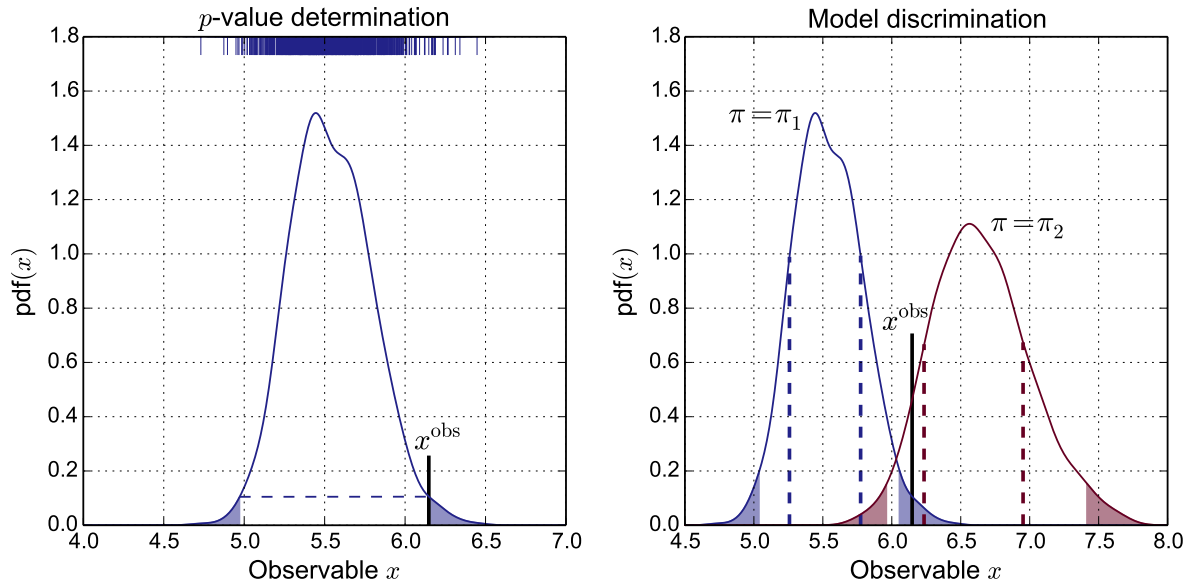
FIGURE 6.9: Examples for the p-value determination and for model selection. The x-axis indicates a one-dimensional data vector, and the y-axis is the PDF, obtained from KDE using the $N = 1000$ realizations. In the left panel, values of realizations are shown as bars in the rug plot at the top of the panel. The shaded area is the corresponding p-value for given observational data x^{obs}. In the right panel, the dashed lines show 1-σ intervals for two models (denoted by π_1 and π_2), and the shaded areas are intervals beyond the 2-σ level. In this example, the model π_1 is excluded at more than 2-σ, whereas the significance of the model π_2 is between 1 and 2-σ.

the "noise" is caused by a relative low number of realizations and/or fluctuation from the Monte Carlo integration. One way to reduce this noise is to reduce the data dimension d. If d is small, on one hand, a low number of realizations can be "sufficient" for reconstructing the PDF; on the other hand, the grid-point evaluation would become tractable. For these reasons, the same analysis is done with some 2D vectors which are obtained by combining components of $\boldsymbol{x}^{\mathrm{abd}}$, and I observe that in some cases the constraint contours are less noisy without deteriorating the constraining power.

Another subtlety from the true likelihood and p-value analysis is the PDF evaluation. While $L_{\mathrm{true}}(\boldsymbol{\pi})$, which is also $P(\boldsymbol{x}^{\mathrm{obs}}|\boldsymbol{\pi})$, is a one-point evaluation at $\boldsymbol{x} = \boldsymbol{x}^{\mathrm{obs}}$, determining p-value requires a more global information, at least about the region where $P(\boldsymbol{x}) < P(\boldsymbol{x}^{\mathrm{obs}})$. This makes evaluating p-values expensive. Furthermore, we should recall that KDE is a biased estimator (Zambom & Dias 2012). What is the bias level compared to the Gaussian likelihood? How to correct it? Those are questions to answer to for further studies. Some further ideas, for example the Voronoi-Delaunay tessellation (see e.g., Schaap 2007; Guio & Achilleos 2009), could be an alternative to the KDE technique.

Summary

I have displayed results of parameter constraints on Ω_{m}-σ_8 provided by our model, by using successively the Gaussian likelihood, the copula likelihood, the true likelihood, and the p-value. The different comparisons have shown that taking the true covariance into account improves the constraining power, and the copula likelihood is a good approximation to the true one.

The results also suggest that the PDF-based data vector, $\boldsymbol{x}^{\mathrm{abd}}$, retains the most cos-

mological information among three candidates. The full CDF-based data vector $\boldsymbol{x}^{\mathrm{pct}}$ also performs well, but could potentially cause serious biases due to our modelling strategy. And the adapted CDF-based data vector $\boldsymbol{x}^{\mathrm{cut}}$ loses information thus carries out larger contours.

The size of the data dimension could affect the performance of the density estimation. In order to find a balance between efficiency and precision, data vector choice, data compression, and density estimation need to be considered jointly, especially for non-parametric estimators.

Recent results from CFHTLenS (Liu J. et al. 2015a) and Stripe-82 (Liu X. et al. 2015b) obtained $\Delta\Sigma_8 \sim 0.1$, about 2–3 times larger than this study. However, I would like to highlight that redshift errors are not taken into account here and that the simulated galaxy density used in this work is much higher. Also, the source redshift has been set to $z_{\mathrm{s}} = 1$, which is higher than the median redshift of both surveys (~ 0.75). All these factors contribute to our smaller error bars.

In a more complete analysis, the number of parameters could be large. The grid-point evaluation will not be available anymore. Instead, sophisticated sampling techniques, such as Markov Chain Monte Carlo (MCMC) or population Monte Carlo (PMC) are usually suggested. With a similar spirit, I am going to present a robust and efficient technique in the next chapter, called approximate Bayesian computation (ABC), which is notably suitable to our FSF model.

Chapter VII

Approximate Bayesian computation

Overview

In this chapter, I will show how to obtain parameter constraints with a method called approximate Bayesian computation. This technique, not so well known by the cosmological community, has a great potential. I will first introduce the concept and describe the philosophy of this method, and then show and comment on the constraint results carried out by a sophisticated iterative algorithm. This chapter corresponds to Sect. 6 of Paper II.

7.1 Introduction to ABC

7.1.1 State of the art

The concept of approximate Bayesian computation (hereafter ABC) has first been mentioned by Rubin (1984) to resolve the problem of the intractable likelihood estimation. For models that have some stochastic characteristics, i.e. from which one can sample, the technique provides an efficient estimation of the posterior. The basic idea of ABC is combining the sampling with an accept-reject criteria in such a way that the process probes the likelihood without evaluating it.

After Rubin (1984), the first ABC-like algorithm was proposed by Tavaré et al. (1997) who studied population genetics. Their algorithm bypasses the likelihood evaluation to estimate the posterior in a particular scenario where values of the data vector are discrete. The generalization for a continuous data space was first proposed by Pritchard et al. (1999). Since then, statistical studies on ABC have become prosperous. Plenty of variations, e.g. Markov Chain Monte Carlo ABC (Marjoram et al. 2003), population Monte Carlo ABC (Beaumont et al. 2009), lazy ABC (Prangle 2016), weighted ABC (Killedar et al. 2015), or Hamiltonian ABC (Meeds et al. 2015), have been proposed.

The technique has been widely utilized, especially in biology-related domains (e.g. Beaumont et al. 2009; Berger et al. 2010; Csilléry et al. 2010; Drovandi & Pettitt 2011). However, applications of ABC for astronomical purposes are still few. These are morphological transformation of galaxies (Cameron & Pettitt 2012), cosmological parameter inference using type Ia supernovae (Weyant et al. 2013; Jennings et al. 2016), constraints of the disk formation of the Milky Way (Robin et al. 2014), cosmological constraints using weak-lensing peak counts

---

Algorithm 1 Discrete ABC

Require:
  observation $\boldsymbol{x}^{\mathrm{obs}}$
  stochastic model $P(\,\cdot\,|\boldsymbol{\pi})$
  number of particles Q
  prior $\mathcal{P}(\,\cdot\,)$

  for $i = 1$ to Q **do**
    repeat
      generate $\boldsymbol{\pi}_i$ from $\mathcal{P}(\,\cdot\,)$
      generate $\boldsymbol{x}$ from $P(\,\cdot\,|\boldsymbol{\pi}_i)$
    until $\boldsymbol{x} = \boldsymbol{x}^{\mathrm{obs}}$ (one-sample test)
  end for

(Paper II, Paper III, Peel et al. 2016), strong lensing properties of galaxy clusters (Killedar et al. 2015), and the constraints on the halo occupation distribution (HOD) parameters (Hahn et al. 2016). Apart from Paper II, three other software packages which allow to perform ABC in a general cosmological context have been released with their respective accompanied papers (Ishida et al. 2015; Akeret et al. 2015; Jennings & Madigan 2016).

7.1.2 The discrete case

How does ABC bypass the likelihood? The philosophy behind it can be easily illustrated in the case of discrete data as follows. Let $\boldsymbol{x}$ be a data vector and $\boldsymbol{\pi}$ a parameter vector. Denote $\mathcal{P}(\,\cdot\,)$ as a prior function, $\mathcal{P}(\,\cdot\,|\boldsymbol{x}^{\mathrm{obs}})$ as the posterior given the observation $\boldsymbol{x}^{\mathrm{obs}}$, and $P(\,\cdot\,|\boldsymbol{\pi})$ as the probability of the model of the data given $\boldsymbol{\pi}$. I would like to remind readers that P is a PDF in the data space and $\mathcal{P}$ in the parameter space. Consider now the accept-reject process from Algo. 1. Each time, we take a parameter candidate $\boldsymbol{\pi}_i$ from the prior, sample a $\boldsymbol{x}$ from the model under $\boldsymbol{\pi}_i$, and accept $\boldsymbol{\pi}_i$ if $\boldsymbol{x}$ and $\boldsymbol{x}^{\mathrm{obs}}$ are identical. In the case of rejection, the parameter should be resampled. We call *particles* all accepted candidates of parameters $\boldsymbol{\pi}_{i=1,\ldots,Q}$ from this algorithm. The number of particles is then Q. Implicitly, each particle is an independent and identically distributed (iid) sample drawn from $\mathcal{P}_{\mathrm{ABC}}(\,\cdot\,|\boldsymbol{x}^{\mathrm{obs}})$, given by

$$
\begin{aligned}
\mathcal{P}_{\mathrm{ABC}}(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}}) &= \sum_{\boldsymbol{x}} \delta_{\boldsymbol{x},\boldsymbol{x}^{\mathrm{obs}}} P(\boldsymbol{x}|\boldsymbol{\pi})\mathcal{P}(\boldsymbol{\pi}) \\
&= P(\boldsymbol{x}^{\mathrm{obs}}|\boldsymbol{\pi})\mathcal{P}(\boldsymbol{\pi}) \\
&= \mathcal{P}(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}}),
\end{aligned}
\tag{7.1}
$$

where the last equality is given by Eq. (6.2) and $\delta_{\boldsymbol{x},\boldsymbol{x}^{\mathrm{obs}}}$ is Kronecker's delta. We can see that the posterior from the ABC process, $\mathcal{P}_{\mathrm{ABC}}(\,\cdot\,|\boldsymbol{x}^{\mathrm{obs}})$, is nothing but the true posterior $\mathcal{P}(\,\cdot\,|\boldsymbol{x}^{\mathrm{obs}})$. Therefore, one is able to reconstruct the posterior without computing the likelihood.

Two important concepts worth being highlighted in Algo. 1. First, if the criterion $\boldsymbol{x} = \boldsymbol{x}^{\mathrm{obs}}$ is not matched, the candidate $\boldsymbol{\pi}_i$ must be dropped and resampled from the prior. The reason is trivial: otherwise, all candidates would eventually be accepted and the ABC process would sample from the prior. Second, it is sufficient to perform a *one-sample test*: a single draw of $\boldsymbol{x}$ from $P(\,\cdot\,|\boldsymbol{\pi}_i)$ for each candidate $\boldsymbol{\pi}_i$. Why is that? It is easy to understand when $\boldsymbol{\pi}$ has also discrete values. In the doubly discrete case, a specific parameter value can be presented by more than one candidate, so asymptotically, having an N-sample test instead

---

Algorithm 2 Continuous ABC

Require:
  requirements of Algo. 1
  summary statistic $s(\cdot)$
  distance $D(\cdot, \cdot)$
  tolerance level ϵ

  for $i = 1$ to Q **do**
    repeat
      generate $\boldsymbol{\pi}_i$ from $\mathcal{P}(\cdot)$
      generate $\boldsymbol{x}$ from $P(\cdot|\boldsymbol{\pi}_i)$
    until $D\left(s(\boldsymbol{x}), s(\boldsymbol{x}^{\mathrm{obs}})\right) \leq \epsilon$ (one-sample test)
  end for

---

is equivalent to setting number of particles N times larger. In a continuous parameter space, the "noise" related to an N-sample test, which is actually the statistical fluctuation related to the likelihood evaluated at a specific point, will be reduced when we reconstruct the global posterior. Recalling that particles are iid samples from the ABC posterior, the true "noise" here is just the same as any reconstruction of an underlying distribution with a limit number of samples.

Another good reason not to use more than one sample is the difficulty to define the accepting criterion. The new criteria might change the probability for accepting $\boldsymbol{\pi}_i$, which is just $\delta_{\boldsymbol{x},\boldsymbol{x}^{\mathrm{obs}}} P(\boldsymbol{x}|\boldsymbol{\pi}_i)$ in the one-sample case. Since this simple form of probability leads to an exact reconstruction of the posterior, the process is already optimal.

7.1.3 The continuous case

How does ABC work in a continuous data space? Since the strict equality of $\boldsymbol{x} = \boldsymbol{x}^{\mathrm{obs}}$ is unlikely to happen in the continuous case, one should introduce a *tolerance level ϵ* below which two data vectors are considered identical. However, real-world data may have very complicated representations and/or a very high dimension, and often need to be reduced to a vector of suitable size. This data compression problem, present in all statistical analyses, is emphasized in the ABC context by adopting the formalism of the *summary statistic s*. Instead of denoting $\boldsymbol{x}$ as a data vector, in the rest of this chapter, $\boldsymbol{x}$ is rather considered as a general data set, and then reduced into a vector $s(\boldsymbol{x})$ via s. Finally, we need a *distance D* to compare in a quantitative way two different data. This could be the Euclidean distance, a covariance-weighted function, or any other customized measure. The distance D is obviously defined in the summary space.

The continuous ABC algorithm is displayed in Algo. 2. Similar to Algo. 1, it generates $\boldsymbol{\pi}_i$ from the prior and does a one-sample test, this time with the criterion $D(s(\boldsymbol{x}), s(\boldsymbol{x}^{\mathrm{obs}})) \leq \epsilon$. In this case, the ABC posterior from which the particles are sampled, denoted as $\mathcal{P}_{\mathrm{ABC},\epsilon}(\cdot|\boldsymbol{x}^{\mathrm{obs}})$, leads to

$$\mathcal{P}_{\mathrm{ABC},\epsilon}(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}}) = A_\epsilon(\boldsymbol{\pi})\mathcal{P}(\boldsymbol{\pi}), \tag{7.2}$$

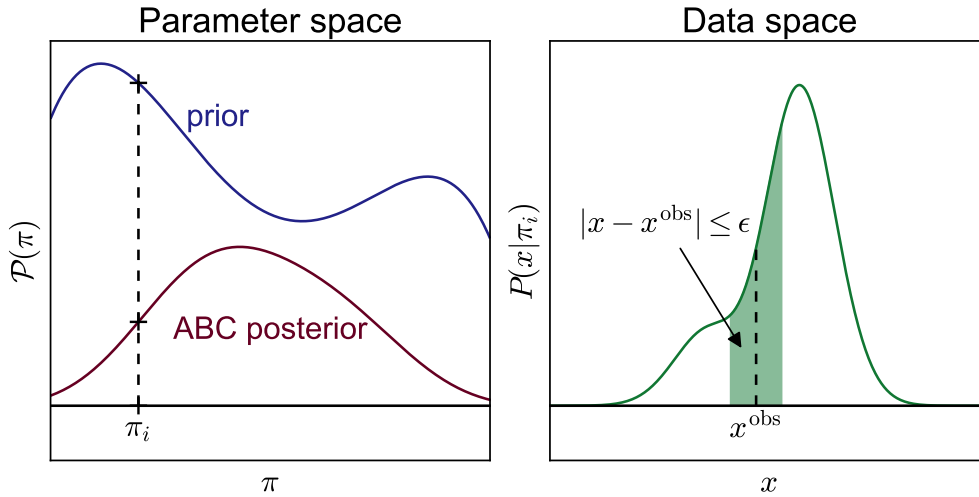where $A_\epsilon(\boldsymbol{\pi})$ is the probability that a candidate $\boldsymbol{\pi}$ passes the one-sample test within the

FIGURE 7.1: Illustration of the ABC posterior. Left panel: the parameter space with the prior in blue and the ABC posterior in dark red. Right panel: the data space with the PDF of x under a specific parameter π_i, in green. Here, $|x - x^{\text{obs}}|$ replaces $D(s(\boldsymbol{x}), s(\boldsymbol{x}^{\text{obs}}))$ to illustrate the difference between x and x^{obs}. If the total integral under the green curve is one, then the green area is the accepting probability, which is also the ratio of the cross on the red curse to the cross on the blue curve.

error ϵ:

$$A_\epsilon(\boldsymbol{\pi}) \equiv \int \mathrm{d}\boldsymbol{x} \; \mathbb{1}_{D(s(\boldsymbol{x}), s(\boldsymbol{x}^{\text{obs}})) \leq \epsilon}(\boldsymbol{x}) P(\boldsymbol{x}|\boldsymbol{\pi}). \tag{7.3}$$

The Kronecker delta from Eq. (7.1) has been replaced with the indicator function $\mathbb{1}$ of the set of points $\boldsymbol{x}$ that satisfy the tolerance criterion. Eqs. (7.2) and (7.3) are visualized in Fig. 7.1, where $\mathcal{P}(\boldsymbol{\pi})$ is represented by the blue curve, $\mathcal{P}_{\text{ABC},\epsilon}(\boldsymbol{\pi}|\boldsymbol{x}^{\text{obs}})$ the red curve, and $P(\boldsymbol{x}|\boldsymbol{\pi})$ the green curve, so that the ratio of the red curve to the blue curve equals to the ratio of the green area to the total integral. From Fig. 7.1, it is straightforward to verify that

$$\lim_{\epsilon \to 0} \frac{A_\epsilon(\boldsymbol{\pi})}{2\epsilon} = P(\boldsymbol{x}^{\text{obs}}|\boldsymbol{\pi}) = \mathcal{L}(\boldsymbol{\pi}|\boldsymbol{x}^{\text{obs}}). \tag{7.4}$$

Therefore, the basic assumption of ABC is to take a small value of ϵ and to consider the ABC posterior as a good approximation of the underlying one, such that

$$\mathcal{P}_{\text{ABC},\epsilon}(\boldsymbol{\pi}|\boldsymbol{x}^{\text{obs}}) \approx \mathcal{P}(\boldsymbol{\pi}|\boldsymbol{x}^{\text{obs}}). \tag{7.5}$$

As a result, the error can be separated into two parts: one from the approximation (7.5) and the other from the estimation of the desired integral (7.2). For the latter, one-sample tests of $A_\epsilon(\boldsymbol{\pi})$ correspond to a Monte Carlo estimation of the ABC posterior, which is unbiased. This ensures the use of the one-sample test.

At the end of the day, ABC can be summarized as a process which, *instead of evaluating the likelihood, samples directly from a posterior which is considered close to the true one; and by reconstructing from samples, one obtains an estimate of the true posterior.*

7.2 Parameter constraints with the PMC ABC algorithm

7.2.1 Choice of the requirements

In Algo. 2, some of the requirements to be chosen by the user can play an important role to the result. First of all, choosing a summary statistic is not obvious. On the one hand, if the considered data vector is too large, the analysis will be slow and inefficient; on the other hand, if data compression is too violent, then we might lose useful information and possibly create biases. A *sufficient* statistic which contains all useful information might be difficult to define. Second, similar to the summary statistic, the choice of the distance function can also be very various. Finally, what about the tolerance level? If ϵ is large, Eq. (7.3) becomes a bad estimate. If ϵ is too small, $A_\epsilon(\boldsymbol{\pi})$ is close to 0 and sampling becomes extremely difficult and inefficient. How to find a trade-off?

A study about the impact of the various choices necessary for ABC has been done by McKinley et al. (2009). The authors found that (1) doing an N-sample test instead of a single sample does not seem to improve the posterior estimation (similar conclusion are found by Bornn et al. 2014; (2) the specific choice of the tolerance level does not seem to be important; and (3) the choice of the summary statistic and the distance is crucial. Therefore, exploring a sufficient summary statistic to represent the whole data set becomes an essential concern for the ABC technique. There exist several methods to select ideal summaries and to compare between different ones, which can be divided into three classes: best subset selection, projection techniques, and regularization approaches (see e.g. Blum et al. 2013). However, users still need to propose a collection of summary candidates. Usually, motivated by computational efficiency, the summary statistic has a low dimension and a simple form in practice. It seems that a simple summary can still produce reliable constraints (Weyant et al. 2013).

As far as the tolerance level is concerned, one of the solutions is population Monte Carlo (PMC, Iba 2001; for applications to cosmology, see Wraith et al. 2009). PMC is an iterative importance sampling process. Its combination with ABC has first been proposed by Beaumont et al. (2009). Meanwhile, based on the theoretical work of Del Moral et al. (2006), Sisson et al. (2007) offered a sequential Monte Carlo (SMC) sampler for application to ABC. After the correction of a bias published by Sisson et al. (2009), this SMC ABC approach is basically identical to the PMC ABC. For this reason, both names can be found in the literature. This iterative solution is the algorithm which has been used for this study.

7.2.2 PMC ABC algorithm

The basic idea of PMC is to generate samples from a *proposal function* which depends on the previous iteration. In the case of the first iteration, the proposal is just the prior. Given a decreasing sequence of $T + 1$ tolerance levels $\epsilon^{(0)} \geq \epsilon^{(1)} \geq \ldots \geq \epsilon^{(T)}$, we start ABC by sampling candidate parameters from the prior. We attribute an equal weight to each particle, and compute the covariance matrix. Then, for all later iterations, candidates are taken from a normal law centered on a particle, randomly selected with regard to the weights. The normal law uses the covariance matrix from the previous iteration. At the end of each iteration, the weights are updated using a transition kernel $W_{\boldsymbol{C}}$, where $\boldsymbol{C}$ is an input covariance matrix (weighted) since the kernel is usually Gaussian. The covariance matrix is also updated by taking new weights into account. The samples from the final iteration are considered as the result from the PMC ABC algorithm, presented in Algo. 3.

Algorithm 3 Population Monte Carlo ABC

Require:
  requirements of Algo. 1
  summary statistic $s(\,\cdot\,)$
  distance $D(\,\cdot\,,\,\cdot\,)$
  transition kernel $W_{\boldsymbol{C}}(\,\cdot\,)$
  decreasing sequence of tolerance levels $\epsilon^{(0)} \geq \epsilon^{(1)} \geq \ldots \geq \epsilon^{(T)}$

  set $t = 0$
  for $i = 1$ to Q **do**
    repeat
      generate $\boldsymbol{\pi}_i^{(0)}$ from $\rho(\,\cdot\,)$ and $\boldsymbol{x}$ from $P\left(\,\cdot\,|\boldsymbol{\pi}_i^{(0)}\right)$
    until $D\left(s(\boldsymbol{x}), s(\boldsymbol{x}^{\mathrm{obs}})\right) \leq \epsilon^{(0)}$
    set $\omega_i^{(0)} = 1/Q$
  end for
  set $\boldsymbol{C}^{(0)} = \mathrm{cov}\left(\boldsymbol{\pi}_i^{(0)}, \omega_i^{(0)}\right)$

  for $t = 1$ to T **do**
    for $i = 1$ to Q **do**
      repeat
        generate j from $\{1, \ldots, Q\}$ with weights $\left\{\omega_1^{(t-1)}, \ldots, \omega_Q^{(t-1)}\right\}$
        generate $\boldsymbol{\pi}_i^{(t)}$ from $\mathcal{N}\left(\boldsymbol{\pi}_j^{(t-1)}, \boldsymbol{C}^{(t-1)}\right)$ and $\boldsymbol{x}$ from $P\left(\,\cdot\,|\boldsymbol{\pi}_i^{(t)}\right)$
      until $D\left(s(\boldsymbol{x}), s(\boldsymbol{x}^{\mathrm{obs}})\right) \leq \epsilon^{(t)}$
      set $\omega_i^{(t)} \propto \dfrac{\mathcal{P}\left(\boldsymbol{\pi}_i^{(t)}\right)}{\sum_{j=1}^{Q} \omega_j^{(t-1)} W_{\boldsymbol{C}^{(t-1)}}\left(\boldsymbol{\pi}_i^{(t)} - \boldsymbol{\pi}_j^{(t-1)}\right)}$
    end for
    set $\boldsymbol{C}^{(t)} = \mathrm{cov}\left(\boldsymbol{\pi}_i^{(t)}, \omega_i^{(t)}\right)$
  end for

In this way, we can actually write the analytical expression for the ABC posterior. The proposal function described above is actually a Gaussian smoothing of the previous posterior. Therefore, the proposal function for the t-th iteration $\widetilde{\mathcal{P}}_t(\boldsymbol{\pi})$ is given by

$$\widetilde{\mathcal{P}}_0(\boldsymbol{\pi}) = \mathcal{P}(\boldsymbol{\pi}), \tag{7.6}$$

$$\widetilde{\mathcal{P}}_t(\boldsymbol{\pi}) = \int \mathrm{d}\boldsymbol{\pi}' \; W_{\boldsymbol{C}^{(t-1)}}(\boldsymbol{\pi} - \boldsymbol{\pi}')\omega_{t-1}(\boldsymbol{\pi}')\mathcal{P}_{t-1}(\boldsymbol{\pi}'|\boldsymbol{x}^{\mathrm{obs}}), \quad \text{for } 1 \leq t \leq T, \tag{7.7}$$

where $\mathcal{P}_t(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}})$ is denoted as the ABC posterior after the t-th iteration, $\omega_t(\boldsymbol{\pi})$ represents the weight function, and $\boldsymbol{C}^{(t)}$ is the weighted covariance matrix derived from $\mathcal{P}_t(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}})$ and $\omega_t(\boldsymbol{\pi})$. This weight function itself depends on the posterior of the previous iteration, so the proposal $\widetilde{\mathcal{P}}_t(\boldsymbol{\pi})$ involves $\omega_{t-1}(\boldsymbol{\pi})$ and $\mathcal{P}_{t-2}(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}})$. Note that in practice, posteriors, proposals, weights, and covariances are computed from discrete samples; here however, we are interested in the expectation values of the sampling process, so all quantities are computed from continuous functions. Considering Eqs. (7.6) and (7.7), the t-th posterior can then be

---

Algorithm 4 Modified PMC ABC used in this work

Require:
  requirements of Algo. 1
  summary statistic $s(\,\cdot\,)$
  distance $D(\,\cdot\,,\,\cdot\,)$
  transition kernel $W_{\boldsymbol{C}}(\,\cdot\,)$
  shutoff parameter r_{stop}

  set $t = 0$
  for $i = 1$ to Q **do**
    generate $\boldsymbol{\pi}_i^{(0)}$ from $\mathcal{P}(\,\cdot\,)$ and $\boldsymbol{x}$ from $P\left(\,\cdot\,|\boldsymbol{\pi}_i^{(0)}\right)$
    set $\delta_i^{(0)} = D\left(s(\boldsymbol{x}), s(\boldsymbol{x}^{\mathrm{obs}})\right)$ and $\omega_i^{(0)} = 1/Q$
  end for
  set $\epsilon^{(1)} = \mathrm{median}\left(\delta_i^{(0)}\right)$ and $\boldsymbol{C}^{(0)} = \mathrm{cov}\left(\boldsymbol{\pi}_i^{(0)}, \omega_i^{(0)}\right)$

  while success rate $\geq r_{\mathrm{stop}}$ **do**
    $t \leftarrow t + 1$
    for $i = 1$ to Q **do**
      repeat
        generate j from $\{1, \ldots, Q\}$ with weights $\left\{\omega_1^{(t-1)}, \ldots, \omega_Q^{(t-1)}\right\}$
        generate $\boldsymbol{\pi}_i^{(t)}$ from $\mathcal{N}\left(\boldsymbol{\pi}_j^{(t-1)}, \boldsymbol{C}^{(t-1)}\right)$ and $\boldsymbol{x}$ from $P\left(\,\cdot\,|\boldsymbol{\pi}_i^{(t)}\right)$
        set $\delta_i^{(t)} = D\left(s(\boldsymbol{x}), s(\boldsymbol{x}^{\mathrm{obs}})\right)$
      until $\delta_i^{(t)} \leq \epsilon^{(t)}$
      set $\omega_i^{(t)} \propto \dfrac{\mathcal{P}\left(\boldsymbol{\pi}_i^{(t)}\right)}{\sum_{j=1}^{Q} \omega_j^{(t-1)} W_{\boldsymbol{C}^{(t-1)}}\left(\boldsymbol{\pi}_i^{(t)} - \boldsymbol{\pi}_j^{(t-1)}\right)}$
    end for
    set $\epsilon^{(t+1)} = \mathrm{median}\left(\delta_i^{(t)}\right)$ and $\boldsymbol{C}^{(t)} = \mathrm{cov}\left(\boldsymbol{\pi}_i^{(t)}, \omega_i^{(t)}\right)$
  end while

---

established via

$$A_{\epsilon_t}(\boldsymbol{\pi}) = \int \mathrm{d}\boldsymbol{x} \; \mathbb{1}_{D(s(\boldsymbol{x}), s(\boldsymbol{x}^{\mathrm{obs}})) \leq \epsilon_t}(\boldsymbol{x}) P(\boldsymbol{x}|\boldsymbol{\pi}), \tag{7.8}$$

$$\mathcal{P}_t(\boldsymbol{\pi}|\boldsymbol{x}^{\mathrm{obs}}) = A_{\epsilon_t}(\boldsymbol{\pi}) \widetilde{\mathcal{P}}_t(\boldsymbol{\pi}). \tag{7.9}$$

The expression is more difficult to interpret now, especially for Eq. (7.7) where the weight function is involved. Qualitatively, each A_{ϵ_t} sharpens the distribution on points where the likelihood is high; and the kernel $W_{\boldsymbol{C}}$ tends to smooth the posterior. As ϵ_t decreases, Bayesian inference tends to dominate over smoothing, thereby yields a converging result.

    Instead of selecting an arbitrary tolerance sequence, I propose a modified version of the PMC ABC algorithm by introducing a *shutoff parameter* r_{stop} (Algo. 4). The algorithm continues until the success rate $r^{(t)}$ of the one-sample test is smaller than r_{stop}. And, inspired by Weyant et al. (2013), each $\epsilon^{(t)}$ is determined by the median of the differences to the observation for all particles of the previous iteration. I also set $\epsilon^{(0)}$ which accepts everything from the prior at $t = 0$. This is the ABC algorithm used in this thesis work.

7.2.3 Settings

In this chapter, parameter constraints using ABC are studied only with two free parameters Ω_m and σ_8. The values of other cosmological parameters are detailed in Sect. 6.1.2. The stochastic model $P(\cdot|\boldsymbol{\pi})$ for Algo. 4 is our FSF peak-count model mentioned in Chap. 5. The settings for the model are identical to the previous chapter, detailed in Table 6.1. The same mock observation is also used.

The study has been done with three summary statistics following Chap. 6. They have been chosen as $s(\boldsymbol{x}) = \boldsymbol{x}^{\text{abd}}$, $\boldsymbol{x}^{\text{pct}}$, and $\boldsymbol{x}^{\text{cut}}$, as defined in Sect. 6.1.3. The corresponding distances are defined as

$$D(\boldsymbol{x}^{\text{type}}, \boldsymbol{y}^{\text{type}}) \equiv \sqrt{\sum_{i=1}^{5} \frac{(x_i^{\text{type}} - y_i^{\text{type}})^2}{C_{ii}}}, \tag{7.10}$$

where `type` is `abd`, `pct`, or `cut`. It is simply a weighted Euclidean distance by neglecting the off-diagonal terms of the covariance matrix. In other words, the cross-correlations between bins are not taken into account. The weighted covariance used in the transition kernel and the normal law is explicitly

$$C_{ij} = \frac{\left(\sum_{k=1}^{Q} \omega_k\right) \cdot \sum_{k=1}^{Q} \omega_k (\pi_{k,i} - \bar{\pi}_i)(\pi_{k,j} - \bar{\pi}_j)}{\left(\sum_{k=1}^{Q} \omega_k\right)^2 - \sum_{k=1}^{Q} \omega_k^2}, \tag{7.11}$$

where $\pi_{k,i}$ stands for the i-th component of the k-th particle and ω_k is associated weight. The inverse is estimated via Eq. (6.14). The prior is flat over a SIM-card-shaped region (see Fig. 7.2). I have set $Q = 250$ and $r_{\text{stop}} = 0.05$. Finally, the Gaussian transition kernel can be explicitly written in Eq. (6.34).

7.2.4 Results

Let us first put the focus on the data vector $\boldsymbol{x}^{\text{abd}}$. For this run, the process stops at the end of the iteration $t = 8$. Figure 7.2 shows ABC posteriors at iterations $0 \leq t \leq 7$. These credible contours are constructed by kernel density estimation (KDE) using Eq. (6.33). Contrary to the proposal function, particle weights are neglected for posterior estimation. When $t = 0$, we can see that the particles (green points) are distributed uniformly in the prior region (white area), as $\epsilon^{(0)} = +\infty$ implies. Because of stochasticity, the input parameter (black star) is even excluded at 1-σ. Gradually, particles move and the usual Ω_m-σ_8 degeneracy direction appears. The credible contours stabilize when $r^{(t)}$ and $\epsilon^{(t)}$ decrease, as indicated in Fig. 7.2.

The stabilization of the contours can be explained by the weight of particles. The left panel of Fig. 7.3 shows particles from the final result with their weights, visualized by both size and color. We observe that isolated particles are more weighted than those in the high-density zone, as constructed by Algo. 4. This fact smooths the posterior, and makes the proposal for the next iteration a flatter function, avoiding undersampling the tails of the posterior and overestimating the constraining power.

The posterior given by the final result is presented on the right panel of Fig. 7.3. This result is compared with the constraint contours obtained with L_{cg} from Eq. (6.9). We can observe a nice agreement between these two constraint techniques, which validates the performance of PMC ABC. However, at both wings of the contour, the ABC contour is less
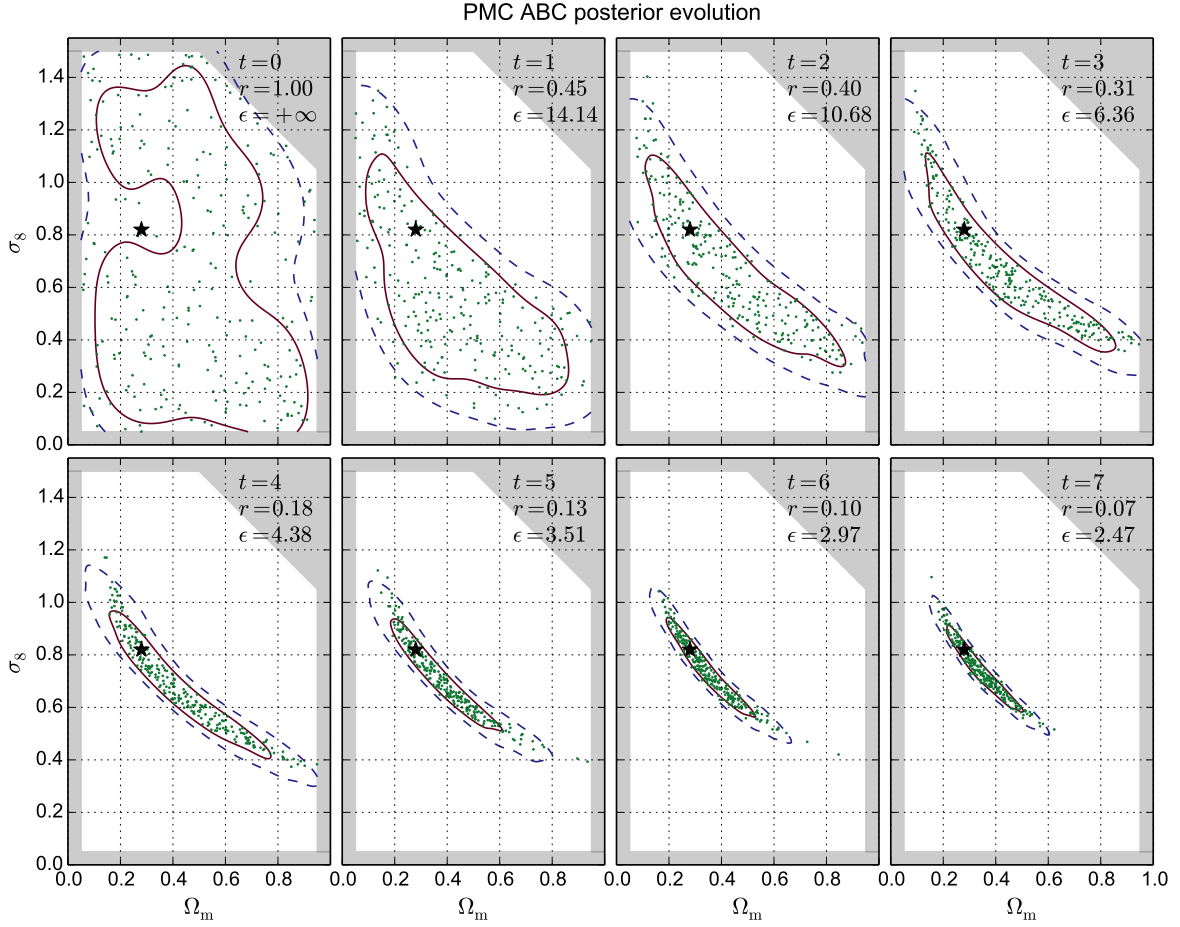
PMC ABC posterior evolution

FIGURE 7.2: Evolution of the Ω_m-σ_8 posterior from the PMC ABC algorithm. Other cosmological parameters are fixed. This figure shows the first eight iterations with their success rate and tolerance level. The prior is presented as the white area, $Q = 250$ particles as green dots, and the input parameter for the observation as the black star. The 1-σ and 2-σ regions are displayed by red solid lines and blue dashed lines, respectively. This run actually extends to $t = 8$ and $t = 9$, which satisfy respectively to $r_{stop} = 0.05$ and 0.03.

curved than the L_{cg} constraint, while by eyes, particles still seem to follow the banana-shaped degeneracy lines. The reason for this discrepancy is KDE. First, KDE smooths and gives systematically a biased estimation. This bias can be more visible at low-probability regions, which is the case of the wings of the contour. Second, the KDE bandwidth matrix is chosen to be diagonal (Eq. (6.35)), so the cross-section of the Gaussian smoothing kernel is an ellipse whose long axis is aligned either with Ω_m-axis or σ_8-axis, not adapted to the degeneracy direction. Nevertheless, with this drawback, KDE still offers a satisfactory result. The values of indicators and best-fits are presented in Tables 6.6 and 6.7. These results are also very consistent with other techniques studied in Chap. 6.

The most important advantage of ABC is efficiency. Start with estimating the time cost in terms of the number of model realizations. Let N be the number of realizations under an input parameter set for estimation of the covariance matrix (for likelihoods) or the PDF (for copula, the true likelihood, and p-value analysis). For parameter constraints by grid-point evaluation or MCMC, the total time cost is $Q \times N$ where Q is the number of parameter sets that are tested. For PMC with likelihoods, we need to add the number of iterations T, so that the time cost becomes $T \times Q \times N$. For instance, for analyses of Chap. 6, $Q = 7821$ and
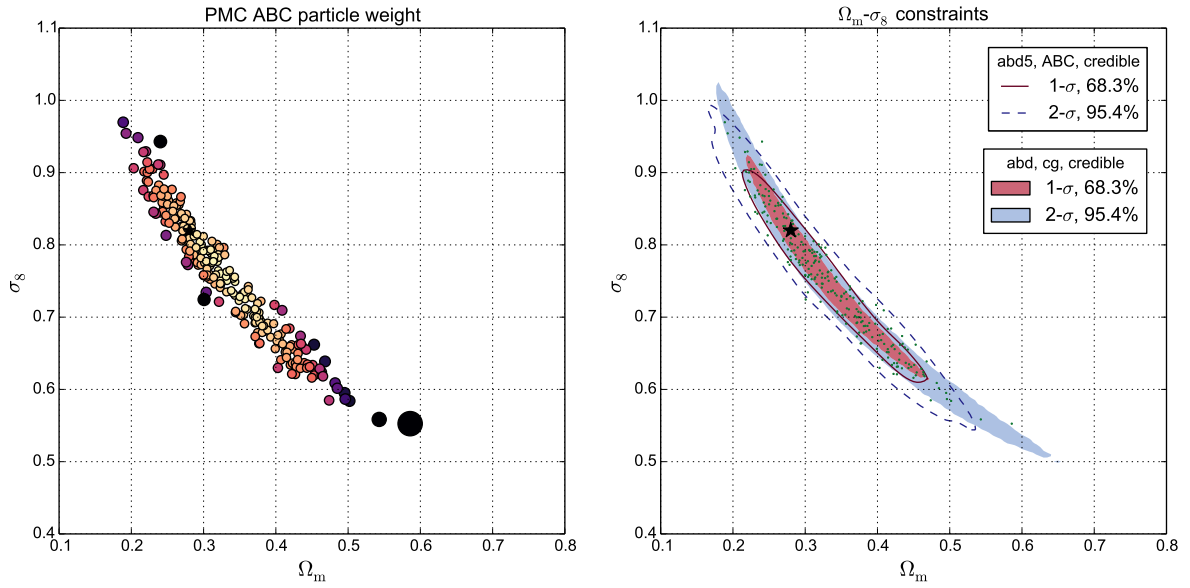
FIGURE 7.3: Results from PMC ABC with $\boldsymbol{x}^{\mathrm{abd}}$. Left panel: particles at $t = 8$ with their weights, illustrated in both size and color. Right panel: comparison between constraints from L_{cg} and ABC; the legend is similar to Fig. 7.2.

$N = 1000$.

What is then the time cost for PMC ABC? In this case, the number of realizations varies between iterations. It is the total number of candidates times one, becuase of the one-sample test. The number of candidates for the t-th iteration is $Q/r^{(t)}$. Thus, the total cost is $\left(\sum_t 1/r^{(t)} \right) \times Q \times 1$. For the run with $\boldsymbol{x}^{\mathrm{abd}}$, $Q = 250$ and $\sum_t 1/r^{(t)} \approx 69$ (for $0 \leq t \leq 8$). So, we end up with 7821×1000 for analyses of Chap. 6 and $69 \times 250 \times 1$ for PMC ABC. The difference is about two orders of magnitude. If we set a more strict shutoff value $r_{\mathrm{stop}} = 0.03$ for PMC ABC (in this case, $\sum_t 1/r^{(t)} \approx 102$ for the same run, $0 \leq t \leq 9$), then the difference is still neatly an order of magnitude. This is more likely the gain of the PMC ABC algorithm for future studies which seek to constrain parameters in a high-dimensional space, where grid-point evaluation is not feasible and Q will take between 10,000 and 100,000 for both MCMC and PMC.

The results for runs with $\boldsymbol{x}^{\mathrm{pct}}$ and $\boldsymbol{x}^{\mathrm{cut}}$ are shown in Fig. 7.4, Table 6.6, and Table 6.7. The conclusion taken from these results is similar to the interpretations above.

Summary

In this chapter, I present a new parameter constraining method called approximate Bayesian computation (ABC). It bypasses likelihood evaluation and generates samples directly from an approximate posterior close to the true one. By reconstructing from samples, one obtains an estimate of the true posterior.

The ABC process requires a stochastic model. The result may depend on the metric which defines the difference between two data sets. This metric involves a distance and the notion of summary statistic which can be interpreted as defining a data vector.

A fashionable way to perform ABC is to combine it with population Monte Carlo (PMC). Thanks to its iterative process, the PMC ABC algorithm decreases the tolerance level and
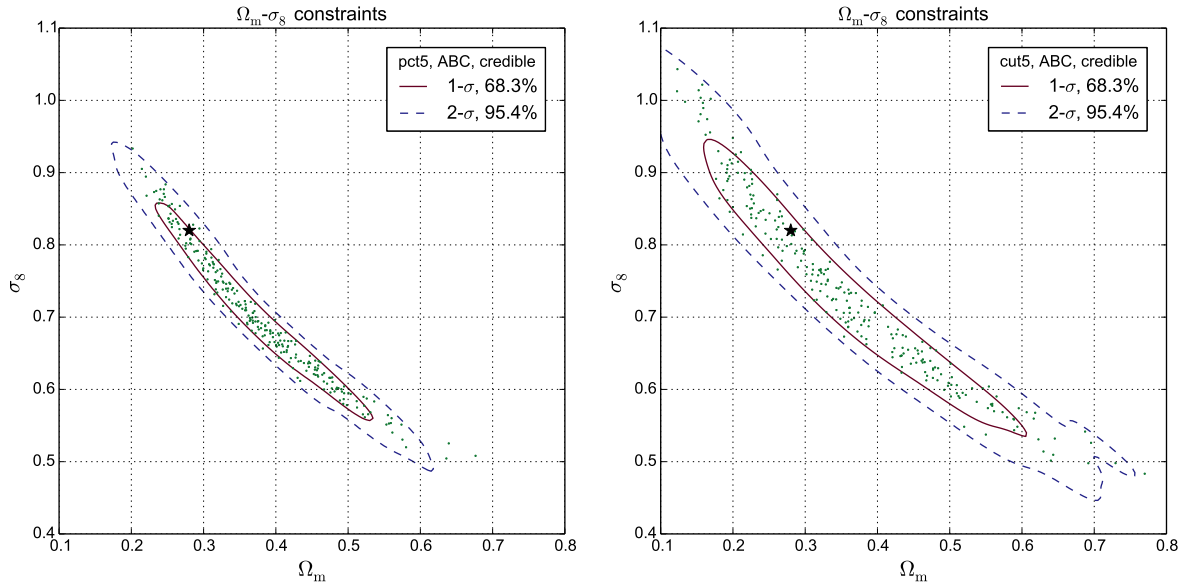
FIGURE 7.4: Credible regions from PMC ABC with $\boldsymbol{x}^{\mathrm{pct}}$ and $\boldsymbol{x}^{\mathrm{cut}}$. These contours correspond respectively to $t = 9$ and $t = 7$. The legend is similar to Fig. 7.2. The wings of contours are also less curved in these cases.

refines the result. In this work, I proposed a modified version with shutoff parameter to decrease the number of arbitrary choices.

The result from PMC ABC agrees well with those from other constraining methods. Only a small number of particles is required. In terms of computational costs, it is two orders of magnitude lower than classical constraining methods using likelihoods. Robust and efficient, PMC ABC seems to be the most adapted solution for stochastic processes, including our model and N-body simulations.

Chapters 5, 6, and 7 have established together a "pipeline" from theoretical inputs to parameter constraints. Before applying this pipeline to the real data, I will take a closer look on filtering techniques in the next chapter.

Chapter VIII

Filtering technique comparisons

Overview

This chapter addresses two questions: the best filter choice and extraction of multiscale information. First, I will argue why some optimal filters proposed in the literature might not be necessarily optimal for cosmological purposes. An alternative standard will be established. Then, several filters, linear or nonlinear, will be compared. Two strategies for dealing with multiscale information will be introduced and tested. All analyses are processed with a realistic version of the fast model, with two parameter constraints methods: the copula likelihood and ABC. This chapter corresponds to Paper III.

8.1   Optimality of the filtering techniques

8.1.1   What are optimal filters optimal for?

In weak lensing, galaxy shape noise usually dominates the lensing signal. Filters are then applied to observations to enhance the peak S/N. Among different filtering methods, some optimization studies, such as Maturi et al. (2005) and Hennawi & Spergel (2005), have been done for peak detections. What objectives are these filters optimal for? They actually refer to different meanings of optimum. On the one hand, Maturi et al. (2005) modelled large-scale structures as "noise" with respect to clusters. Following this reasoning, given a halo density profile on a given scale, the ideal shape for the smoothing kernel can be obtained. On the other hand, Hennawi & Spergel (2005) constructed a *tomographic matched filter* algorithm. Given a kernel shape, this algorithm was able to determine the most probable position and redshift of presumed clusters.

Recall the vocabulary defined in Sect. 4.5: *cluster-oriented purposes* means that cluster detection is put forward, even though these detections may be used to probe cosmology later; *cosmology-oriented purposes* seek to constrain directly cosmology, bypassing cluster identification. With these concepts, readers can find out that both methods mentioned above aim for cluster-oriented purposes, and these might not be optimal if we are interested in cosmological constraints.

Why does the optimal cluster detection not imply the maximal cosmological information extraction from WL peaks? First, a peak is not necessarily created by a single halo (Yang

et al. 2011; Liu & Haiman 2016). Optimizing multiple-halo peaks with single-halo detection approaches is counter-effective. Second, large-scale structures could contribute to peak counts and increase cosmological implications from peaks. As we go to regimes with lower S/N, these effects could become non-negligible. Furthermore, cluster-oriented studies usually focus on purity and completeness as performance indicators (Hennawi & Spergel 2005; Pires et al. 2012; Leonard et al. 2014). This duality makes the comparisons between filters ill-defined. Even if we only focus on one single detection method, the choice between high purity and high completeness is still ambiguous. What strategy extracts more cosmological information? What is its link to parameter constraints? These subtle questions remain unsolved.

For peak studies aiming for cosmology-oriented purposes, a more proper way to find the filtering optimality is to examine directly the cosmological results. This means that instead of looking at purity and completeness, we should compare directly the resulting constraints contours (width, size, etc.) between different filters. Under this standard, no filtering technique can claim to be optimal in an "obvious" way — especially when we want to explore multiscale information. The only justification we can give is to compare between some potential candidates, and find out the one which outperforms a necessarily non-exhaustive list of other filters. This idea will be strictly followed in this chapter.

8.1.2 Separated and combined strategies

Compared to the studies in previous chapters, an improvement has been put forward by exploring multiscale information instead of one single filter scale. There exist two strategies to deal with multiple scales, namely the separated and combined strategies.

The *separated strategy* (followed implicitly by Maturi et al. 2005, 2010; see also Liu J. et al. 2015a) applies a series of filters at different scales. This results in a series of filtered weak-lensing maps, and cosmological constraints are derived from the ensemble of these maps. For weak-lensing peaks, we can for example concatenate peak-count histograms to do parameter inference.

The *combined strategy* (followed e.g. by Hennawi & Spergel 2005; Marian et al. 2012), sometimes called mass mapping, yields only one map. This means that the significances from different scales have been compared and summarized into a representative value. Both Hennawi & Spergel (2005) and Marian et al. (2012) use a likelihood approach to choose, for each position, the most adaptive filtering scale, so that only one filtered map is provided, from which we can estimate peak abundance and derive constraints.

Apart from linear filters, there also exist various nonlinear reconstruction techniques. For example, Bartelmann et al. (1996) proposed to minimize the error on shear and magnification together. Other techniques are sparsity-based methods such as MRLens (Starck et al. 2006), FASTLens (Pires et al. 2009a), and Glimpse (Leonard et al. 2014). These approaches aim to map the projected mass through a minimization process. Therefore, the optimality of filters is a complex issue involving several aspects such as linearity, either separated or combined strategy for multiscale information, and the choice of scales. The aim of this chapter is to address this problem and to answer to following questions:

- For a given kernel shape, with the separated strategy, what are the preferable characteristic scales?

- Which can extract more cosmological information, the compensated or non-compensated filters?

- Which can extract more cosmological information, the separated or combined strategy?

- How do nonlinear filters perform?

8.2 Linear filters

8.2.1 Convergence filters and noise level

In this section, I will re-establish the formalism for filtering. Consider a noisy convergence map to filter. Let $\boldsymbol{\theta}$ be angular coordinates. Let θ_{ker} be the size of the filtering kernel and $x = \theta/\theta_{\mathrm{ker}}$. The filter function will always be described by $W(x)$ with the dimensionless position x. A very common choice of W is the Gaussian smoothing kernel:

$$W(x) \propto \exp\left(-x^2\right). \tag{8.1}$$

In principle, we assume that both components of the galaxy intrinsic ellipticity $\epsilon = \epsilon_1 + \mathrm{i}\epsilon_2$ follow the same Gaussian distribution. By noting $\sigma_\epsilon^2 = \sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2$ as the sum of both variances, the noise for the smoothed convergence is also Gaussian. The new variance depends on the kernel shape and the expected galaxy number density n_{gal}, which results in the global noise level (see e.g. Van Waerbeke 2000):

$$\sigma_{\mathrm{noise}}^2 = \frac{\sigma_\epsilon^2}{2n_{\mathrm{gal}}} \cdot \frac{\parallel W \parallel_2^2}{\parallel W \parallel_1^2}, \tag{8.2}$$

where $\parallel W \parallel_p$ stands for the p-norm of W. Equation (8.2) is valid for unnormalized kernels because the normalization factors from the denominator and the numerator cancel each other. For example, if W is a Gaussian with width θ_{ker}, then $\parallel W \parallel_2^2 / \parallel W \parallel_1^2 = 1/2\pi\theta_{\mathrm{ker}}^2$.

Equation 8.2 is the global noise level since n_{gal} is the global density. This is the true noise only if sources are distributed regularly. In realistic conditions, random fluctuations, mask effects, and clustering of source galaxies lead to irregular distributions, which results in a non-constant noise level. To properly take this into account, we can transform norms into discrete sums. Thus, the local noise level is

$$\sigma_{\mathrm{noise}}^2(\boldsymbol{\theta}) = \frac{\sigma_\epsilon^2}{2} \cdot \frac{\sum_i W^2(\boldsymbol{\theta}_i - \boldsymbol{\theta})}{\left(\sum_i |W(\boldsymbol{\theta}_i - \boldsymbol{\theta})|\right)^2}, \tag{8.3}$$

where $\boldsymbol{\theta}_i$ is the position of the i-th galaxy, and i runs over non-masked galaxies under the kernel W. This is nothing but the variance of a weighted sum of independent Gaussian random variables each with variance $\sigma_\epsilon^2/2$.

In this work, the 2D *starlet* function (Starck et al. 2002) is also studied. It is defined as

$$W(x, y) = 4\phi(2x)\phi(2y) - \phi(x)\phi(y), \tag{8.4}$$

where $x = \theta_1/\theta_{\mathrm{ker}}$, $y = \theta_2/\theta_{\mathrm{ker}}$, and ϕ is the B-spline of order 3, given by

$$\phi(x) = \frac{1}{12}\left(|x-2|^3 - 4|x-1|^3 + 6|x|^3 - 4|x+1|^3 + |x+2|^3\right). \tag{8.5}$$

Because of the property the B-spline, the starlet is a compensated function with compact support in $[-2, 2] \times [-2, 2]$. It does not conserve circular symmetry, but its isolines tend to be round (Fig. 8.1). For the starlet, $\parallel W \parallel_1^2 \approx 0.979^2$ and $\parallel W \parallel_2^2 = 5(2/5 + 5/63)^2 - 2(1/3 + 1/5 + 1/21 + 1/48)^2 \approx 0.652^2$ can be solved analytically.
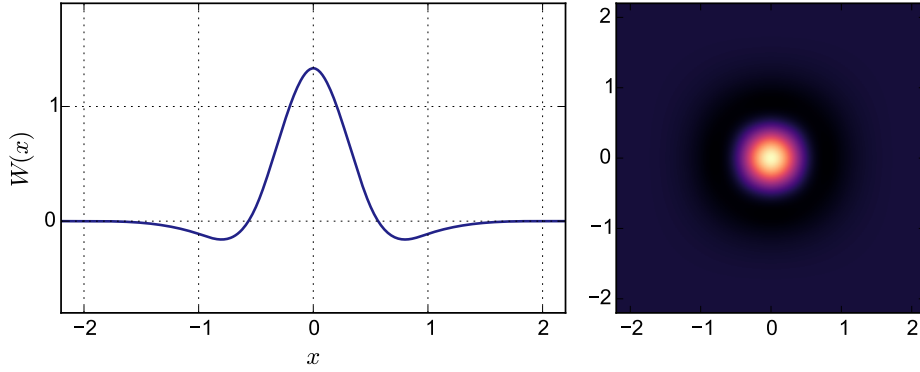
FIGURE 8.1: Left panel: the profile of the 2D starlet. It has a finite support [-2, 2]. Right panel: the bird-eye view of the 2D starlet.

8.2.2 Aperture mass

The *aperture mass* M_{ap} (Kaiser et al. 1994; Schneider 1996) was motivated by the fact that the convergence is not directly observable, and that shear inversion generates a free constant κ_0. To avoid this, one may convolve directly the shear with another filter and expect that this is equivalent to convergence filtering. This is possible only if the filter is circular. In this case, the convolution product is then the aperture mass.

More precisely, we look for all pairs of filters (U, Q) such that (1) U is circularly symmetric, (2) U is a compensated function, and (3) $U * \kappa = Q * \gamma_+$. With these conditions, convolving γ_+ with Q results in a filtered convergence map that is not affected by either κ_0 or inversion systematics.

To satisfy the third condition, Q has to be related to U by

$$Q(\theta) \equiv \frac{2}{\theta^2} \int_0^\theta \mathrm{d}\theta' \; \theta' U(\theta') - U(\theta). \tag{8.6}$$

Equation (8.6) is directly derived from Eq. (4.55). Then, M_{ap} is given by

$$M_{\mathrm{ap}}(\boldsymbol{\theta}) \equiv \int \mathrm{d}^2\boldsymbol{\theta'} \; U(\boldsymbol{\theta})\kappa(\boldsymbol{\theta} - \boldsymbol{\theta'}) = \int \mathrm{d}^2\boldsymbol{\theta'} \; Q(\boldsymbol{\theta})\gamma_{\mathrm{t}}(\boldsymbol{\theta} - \boldsymbol{\theta'}). \tag{8.7}$$

Here, we are particularly interested in the Q function proposed by Schirmer et al. (2004) and Hetterscheidt et al. (2005), given by

$$Q(x) \propto \frac{\tanh(x/x_{\mathrm{c}})}{(x/x_{\mathrm{c}})\left(1 + \exp(a - bx) + \exp(-c + dx)\right)}, \tag{8.8}$$

with $a = 6$, $b = 150$, $c = 47$, $d = 50$ to have a cutoff around $x = 1$. This filter shape has been motivated by the tangential shear pattern generated by NFW halo profiles. Also, we set $x_{\mathrm{c}} = 0.1$ as suggested by Hetterscheidt et al. (2005).

For the noise level of a map filtered with M_{ap}, Eqs. (8.2) and (8.3) are still valid. We only need to substitute W with Q (Schneider 1996).

8.3 A sparsity-based nonlinear filter

In this section, a nonlinear filtering technique using the sparsity of signals is introduced.

8.3.1 What is sparsity?

In signal processing, a signal is sparse in a specific representation if most of the information is contained in only a few coefficients. This means that either only a finite number of coefficients is non zero, or the coefficients decrease fast when rank-ordered.

A straightforward example is the family of sine functions. In real space, sine functions are not sparse. However, they are sparse in Fourier space since they become the Dirac delta functions. More generally, periodic signals are sparse in the Fourier space.

Why is this interesting? Because white noise is not sparse in any representation. Therefore, if the information of the signal can be compressed into a few strong coefficients, it can easily be separated from the noise. This concept of sparsity has been widely used in the signal processing domain for applications such as denoising, inpainting, deconvolution, inverse problems, or other optimization problems (Daubechies et al. 2004; Candes & Tao 2006; Elad & Aharon 2006; Candès et al. 2008; Fadili et al. 2009). Examples can also be found for studying astrophysical signals (Lambert et al. 2006; Pires et al. 2009b; Bourguignon et al. 2011; Carrillo et al. 2012; Bobin et al. 2014; Ngolè Mboula et al. 2015; Lanusse et al. 2016).

8.3.2 Wavelet transform

From the previous section, one can see that the sparsity of a signal depends on its representation basis. In which basis is the weak lensing signal sparse? A promising candidate is the wavelet transform which decomposes the signal into a family of scaled and translated functions. Wavelet functions are all functions ψ which satisfy the *admissibility condition*:

$$\int_0^{+\infty} \frac{\mathrm{d}k}{k} \; |\tilde{\psi}(k)|^2 < +\infty. \tag{8.9}$$

For example, the starlet is one of the wavelet functions. One of the properties implied by this condition is $\int \psi = 0$, which restricts ψ to a compensated function. In other words, one can consider wavelet functions as highly localized functions with a zero mean. Such a function ψ is called the mother wavelet, which can generate a family of daughter wavelets such as

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi \left(\frac{x-b}{a} \right), \tag{8.10}$$

which are scaled and translated versions of the mother ψ.

The wavelet transform (see e.g. Chaps. 2 and 3 of Starck et al. 2002) refers to the decomposition of an input image into several images of the same size each associated to a specific scale. Due to the property of wavelet functions, each resulting image gives the details of the original one at different scales. If we stack all the images, we recover the original signal.

In the peak-count scenario, peaks which are generated by massive clusters are considered as signals. Like clusters, these signals are local point-like features, and therefore have a sparse representation in the wavelet domain. As described in Sect. 8.3.1, white noise is not sparse. So one simple way to reduce the noise is to transform the input image into the wavelet domain, set a relatively high threshold λ, cut out weak coefficients smaller than λ, and reconstruct the clean image by stacking the thresholded images. This introduces non-linearity. In this chapter, we use the 2D starlet function as the mother wavelet, given by Eq. (8.4), which satisfies the admissibility condition.

Kernel	Gaussian				
$\theta_{\rm ker}$ [arcmin]	1.2	2.4	4.8		
$\sigma_{\rm noise}$	0.027	0.014	0.0068		
Kernel	Starlet				
$\theta_{\rm ker}$ [arcmin]	2	4	8	12	16
$\sigma_{\rm noise}$	0.027	0.014	0.0068	0.0045	0.0034
Kernel	$M_{\rm ap}$ tanh				
$\theta_{\rm ker}$ [arcmin]	2.125	4.25	8.5		
$\sigma_{\rm noise}$	0.027	0.014	0.0068		

TABLE 8.1: List of kernel sizes $\theta_{\rm ker}$. These are chosen based on $\sigma_{\rm noise}$ such that the corresponding values are similar. The quantity $\sigma_{\rm noise}$ is computed using Eq. (8.2) with $n_{\rm gal} = 12$ arcmin^{-2} and $\sigma_\epsilon = 0.4$.

8.3.3 The MRLens filter

In this study, the nonlinear filtering technique *MultiResolution tools for gravitational Lensing* (MRLens, Starck et al. 2006) is applied to lensing maps. MRLens is an iterative filtering based on a Bayesian framework that uses a multiscale entropy prior and the false discovery rate (FDR, Benjamini & Hochberg 1995) which allows to derive robust detection levels in wavelet space.

More precisely, MRLens first applies a wavelet transform to a noisy map. The mother wavelet is chosen to be the starlet function. Then, in the wavelet domain, it determines the threshold by FDR. The denoising problem is regularized using a multiscale entropy prior only on the non-significant wavelet coefficients. Readers are welcome to read Starck et al. (2006) for a detailed description of the method.

Note that, whereas Pires et al. (2009a) selected peaks from different scales separately before the final reconstruction, here peaks are counted on the final reconstructed map. Actually, the methodology of Pires et al. (2009a) is close to filtering with a lower cutoff in the histogram defined by FDR, thus similar to starlet filtering. With the vocabulary defined in Sect. 8.1.2, Pires et al. (2009a) followed the separated strategy and here we attempt the combined strategy. This choice provides a comparison between cosmological information extracted with two strategies, by comparing starlet filtering to the MRLens case.

8.4 Methodology

8.4.1 Filter comparisons

Comparing different filtering techniques for peak counts is the subject of this analysis. This comparison contains (1) the Gaussian kernel, (2) the starlet function, (3) the aperture mass with the hyperbolic tangent function, and (4) the nonlinear filtering technique MRLens.

The linear filters are parametrized with a single parameter, which is the size of the kernel $\theta_{\rm ker}$. In order to compare in a pertinent way bewteen multiscale information captured by different filter shapes, two rules are proposed. The first is to choose $\theta_{\rm ker}$ such that the 2-norms have the same value if kernels are normalized (by their respective 1-norms). The reason for this is that if the ratio of the 2-norm to the 1-norm is identical, then the comparison is based
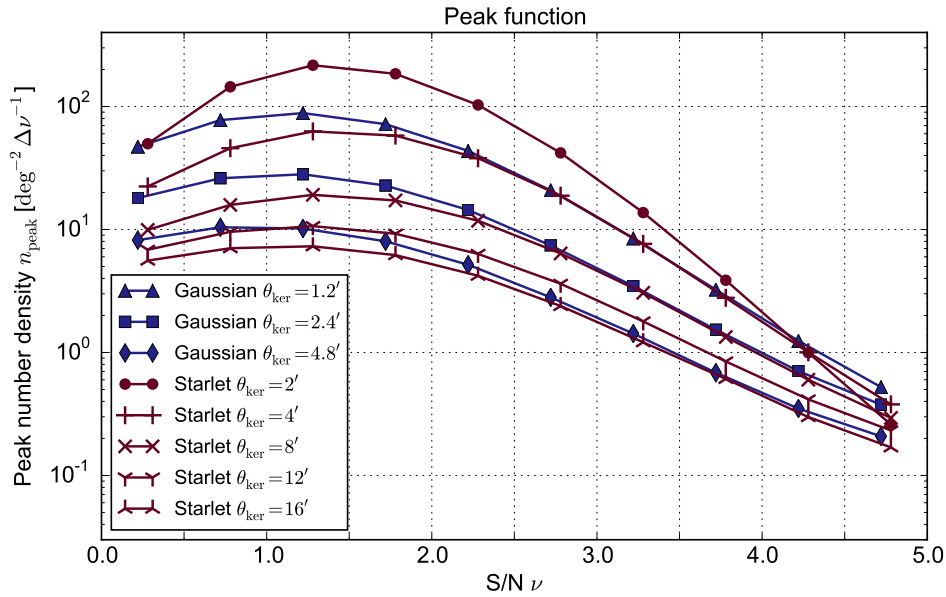
Peak function

FIGURE 8.2: Peak function for different kernel sizes for an input cosmology $(\Omega_{\mathrm{m}}, \sigma_8, w_0^{\mathrm{de}}) = (0.28, 0.82, -0.96)$. The number counts are the mean over 400 realizations of 36 deg². Focusing on the range $2.5 \leq \nu \leq 4.5$, we find that the average number counts of Gaussian filtering with $\theta_{\mathrm{ker}} = 1.2$, 2.4, and 4.8 arcmin correspond respectively to starlet filtering with $\theta_{\mathrm{ker}} = 4$, 8, and 16 arcmin.

on the same global noise level (Eq. 8.2). Table 8.1 shows various values of θ_{ker} that are used in this study and the corresponding σ_{noise} for different linear filters. For the Gaussian filter with $\theta_{\mathrm{ker}} = 1.2$, 2.4, and 4.8 arcmin, the corresponding scales are $\theta_{\mathrm{ker}} = 2$, 4, and 8 arcmin for the starlet function.

The second way is to calculate peak-count histograms, and set θ_{ker} such that peak abundance is similar. Figure 8.2 shows an example for the Gaussian and starlet kernels with θ_{ker} taken from Table 8.1. We can observe that, for Gaussian filtering with $\theta_{\mathrm{ker}} = 1.2$, 2.4, and 4.8 arcmin, the correspondence for starlet filtering based on peak abundance is $\theta_{\mathrm{ker}} = 4$, 8, and 16 arcmin, if we focus on peaks with $2.5 \leq \nu \leq 4.5$.

The MRLens filter handles different scales by the wavelet transform mentioned in Sect. 8.3.2. The most important parameter is the FDR α. It is set to $\alpha = 0.05$ for this analysis.

8.4.2 Settings for the pipeline: from the mass function to peak catalogues

Compared to previous analyses, more realistic observational features are accounted for now. We apply a redshift distribution for source galaxies, include masks, construct the convergence from the reduced shear instead of computing it directly, test different filters, determine the noise level locally, and include the equation of state of dark energy for the constraints. A summary is illustrated by Fig. 8.3. Details are described below.

Fast simulations

For fast simulations, the settings remain similar to previous chapters. Halos are sampled from the model of Jenkins et al. (2001, Eq. 3.81). The sampling range is set to $M = [5 \times 10^{12}, 10^{17}]\ M_\odot/h$. This is done for 60 equal redshift bins from 0 to 3, on a field ade-
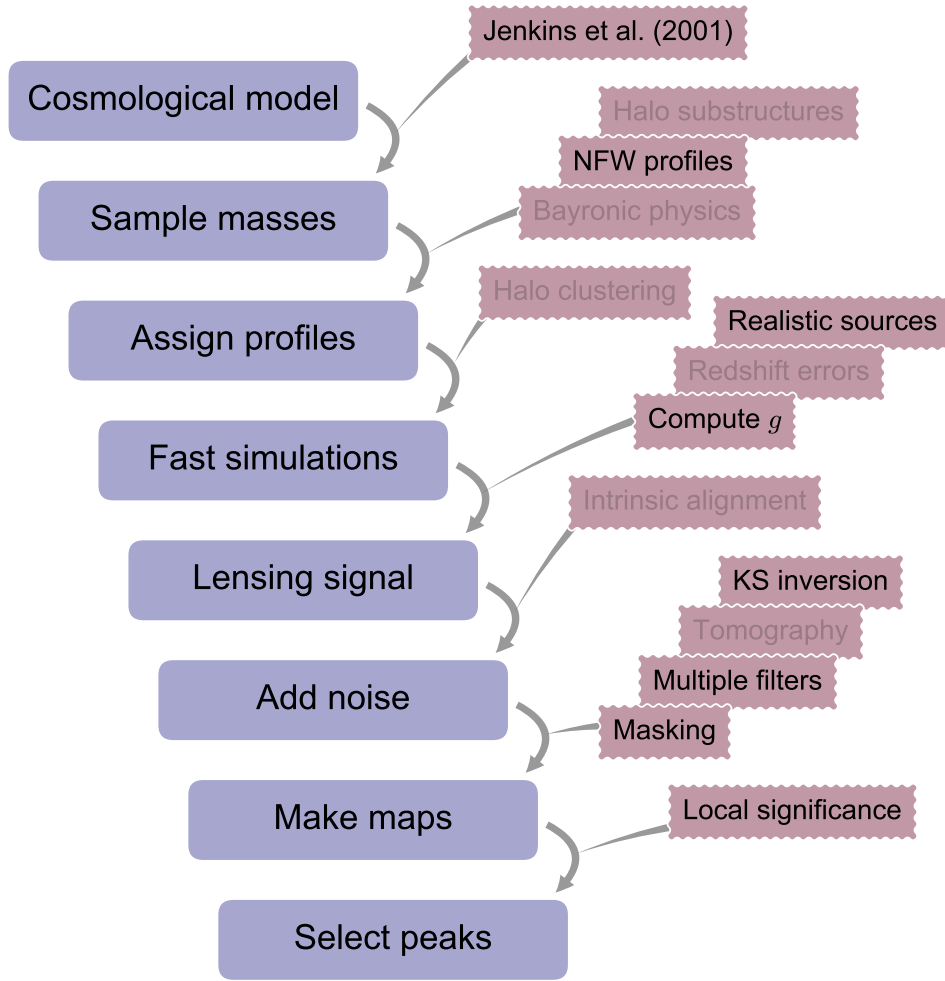
FIGURE 8.3: Diagram illustrating the processing pipeline for the filtering comparison. Effects which are not considered are made faint. A more detailed description for all variations can be found in Sect. 5.4.

quately larger than 36 deg^2 so that border effects are properly eliminated. Halo correlation is neglected. The truncated NFW profiles (Eq. 3.92) are assigned to each halo. The M-c relation is given by Eq. (3.93).

Source catalogues

Instead of sources at a fixed plane, a realistic redshift distribution is considered here. This is assumed to be a gamma distribution following Efstathiou et al. (1991):

$$p(z) = \frac{z^2}{2z_0^3} \exp\left(-\frac{z}{z_0}\right),$$
(8.11)

where $z_0 = 0.5$ is the pivot redshift value. The positions of sources are random, but masking has been taken into account. The masked area of the W1 field of CFHTLenS has been taken as the characteristic mask and is applied to all simulations. The source number density is set to $n_{\rm gal} = 12$ arcmin^{-2}, which corresponds to a CFHTLenS-like survey (Heymans et al. 2012). The intrinsic ellipticity dispersion is $\sigma_\epsilon = 0.4$, which is also close to the CFHTLenS survey (Kilbinger et al. 2013).

Parameter	Symbol	Value
Lower sampling limit	M_{min}	$5 \times 10^{12}\ M_\odot/h$
Upper sampling limit	M_{max}	$10^{17}\ M_\odot/h$
Maximum halo redshift	-	3
Number of halo redshift bins	-	60
NFW inner slope	α	1
M-c relation amplitude	c_0	11
M-c relation power law index	β_{NFW}	0.13
Intrinsic ellipticity dispersion	σ_ϵ	0.4
Galaxy number density	n_{gal}	12 arcmin^{-2}
Pixel size	θ_{pix}	0.8 arcmin
Effective field area	-	36 deg^2
Threshold for filling factor	λ	0.5

TABLE 8.2: List of parameter values adopted in the study of this chapter.

Ray-tracing

For each galaxy, we compute κ_{proj} and γ_{proj} following

$$\kappa_{\mathrm{proj}}(\boldsymbol{\theta}) = \sum \kappa_{\mathrm{halo}}(\boldsymbol{\theta}, w_\ell, w_{\mathrm{s}}) \quad \text{and} \quad \gamma_{\mathrm{proj}}(\boldsymbol{\theta}) = \sum \gamma_{\mathrm{halo}}(\boldsymbol{\theta}, w_\ell, w_{\mathrm{s}}), \qquad (8.12)$$

where the sums run over all lens-source pairs and κ_{halo} and γ_{halo} are given respectively by Eqs. (4.63) and (4.64). The observed ellipticity is computed as $\epsilon^{(\mathrm{o})} = (\epsilon^{(\mathrm{s})} + g_{\mathrm{proj}})/(1 + g^*_{\mathrm{proj}}\epsilon^{(\mathrm{s})})$, where $g_{\mathrm{proj}} \equiv \gamma_{\mathrm{proj}}/(1 - (\kappa_{\mathrm{proj}} - \overline{\kappa}_{\mathrm{proj}}))$ is the reduced shear and $\epsilon^{(\mathrm{s})}$ is the intrinsic ellipticity, taken from a random generator. The subtraction from κ_{proj} of its mean over the field is required by the fact that it is always positive and can not be considered as the true convergence, as already evoked earlier.

Map making

Except for the aperture mass, galaxies are first binned into map pixels for the reason of efficiency. The mean of $\epsilon^{(\mathrm{o})}$ is taken as the pixel's value. The pixel size is set to 0.8 arcmin. This results in regularly spaced data so that the algorithm can be accelerated. Then, the linear KS inversion is applied before filtering. The iterative correction for the reduced shear is not used. By applying exactly the same processing to both observation and model prediction, we expect the systematics related to inversion (e.g. boundary effects, missing data, and negative mass density) to be similar so that the comparison is unbiased.

For the aperture mass, the pixel's value is evaluated by convolving directly the lensing catalogue with the Q filter (Eq. 8.8), successively placed at the center of each pixel (see also Marian et al. 2012; Martinet et al. 2015). The choice of filter sizes is detailed in Sect. 8.4.1.

Peak selection

For linear filters (the Gaussian, the starlet, the aperture mass), peaks are now selected based on their local noise level, determined by Eq. (8.3). In this case, the S/N of a peak is a local

maximum with

$$\nu(\boldsymbol{\theta}) \equiv \begin{cases} (\kappa * W)(\boldsymbol{\theta})/\sigma_{\mathrm{noise}}(\boldsymbol{\theta}) & \text{if Gaussian or starlet,} \\ M_{\mathrm{ap}}(\boldsymbol{\theta})/\sigma_{\mathrm{noise}}(\boldsymbol{\theta}) & \text{if aperture mass.} \end{cases} \tag{8.13}$$

For the nonlinear filter, the notion of noise level does not easily apply. Actually, to determine the significance of a rare event from any distribution, instead of using the empirical standard deviation, it is more rigorous to obtain first the p-value and find how much σ this value is associated with if the distribution was Gaussian. However, even if we compute the standard deviation instead, this process is still too expensive computationally for our purpose. Therefore, we simply select peaks on κ.

Because of masks, only pixels satisfying a specific condition are examined. This criterion is based on the filling factor $f(\boldsymbol{\theta})$ (Van Waerbeke et al. 2013; Liu X. et al. 2015b). A local maximum is selected as a peak only if $f(\boldsymbol{\theta}) \geq \lambda \bar{f}$, where $\bar{f}$ is the mean of f over the map. We set $\lambda = 0.5$. For analyses using binning, the filling factor is simply defined as the number of galaxies $N(\boldsymbol{\theta})$ inside the pixel at $\boldsymbol{\theta}$. For the aperture mass, it is the Q-weighted sum of the number counts. In other words,

$$f(\boldsymbol{\theta}) \equiv \begin{cases} N(\boldsymbol{\theta}) & \text{if galaxies are binned,} \\ \sum_i Q(\boldsymbol{\theta}_i) & \text{for the aperture mass,} \end{cases} \tag{8.14}$$

where $\boldsymbol{\theta}_i$ is the position of the i-th galaxy. Also, we only count pixels in the inner area to avoid border effects. The size of this effective field is 36 deg^2.

8.4.3 Data vector definitions

The data vector $\boldsymbol{x}$, for linear filters, is defined as the concatenation of S/N histograms from various scales. In Chap. 6, we have seen that the number counts from histograms are the most appropriate form to derive cosmological information from peak counts. After testing several values of ν_{min}, we only keep peaks above $\nu_{\mathrm{min}} = 1$ for each kernel size. This choice maximizes the figure of merit of parameter constraints. Thus, the statement from Yang et al. (2013) that ignoring peaks with $\nu \leq 3$ corresponds to a loss of cosmological information is reconfirmed. For each scale, peaks are binned with width of $\Delta\nu = 0.5$ up to $\nu = 5$, and the last bin is $[5, +\infty[$.

For the nonlinear filter, peaks are binned directly by their κ values into $[0.02, 0.03, 0.04, 0.06, 0.10, 0.16, +\infty[$. This configuration is chosen such that the average count per bin is large enough to assume a Gaussian fluctuation.

8.4.4 Sampling in the parameter space

The free parameters to constrain in this study are $(\Omega_{\mathrm{m}}, \sigma_8, w_0^{\mathrm{de}})$. The values of other cosmological parameters are $h = 0.78$, $\Omega_{\mathrm{b}} = 0.047$, and $n_{\mathrm{s}} = 0.95$. The Universe is assumed to be flat.

The mock observation is generated by a realization of our model, using a particular set $(\Omega_{\mathrm{m}}, \sigma_8, w_0^{\mathrm{de}}) = (0.28, 0.82, -0.96)$ as input parameters. In this way, we only focus on the precision of our model. Simulations runs are proceeded in two different ways. The first one consists of interpolating the likelihood, from which we draw credible regions from Bayesian inference, and the second is ABC.
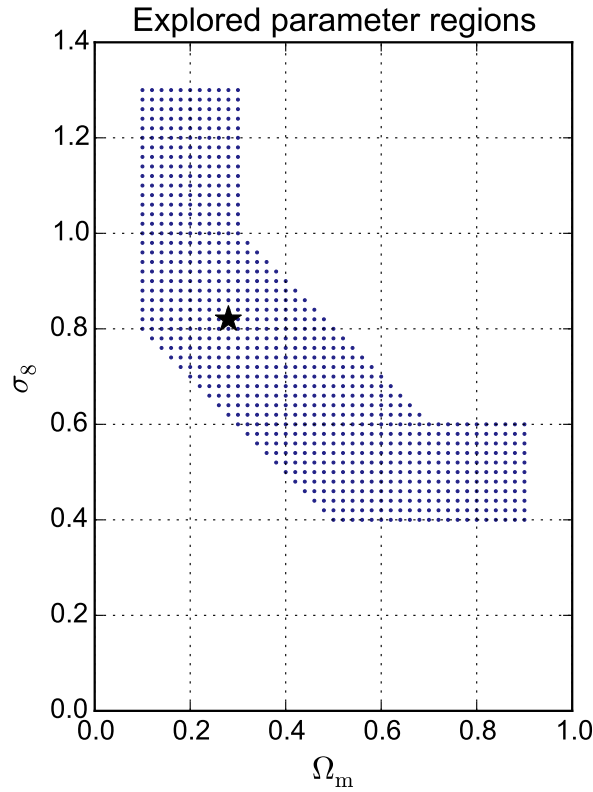
FIGURE 8.4: Distribution of evaluated parameter points on the Ω_m-σ_8 plane. This figure can be considered as a slice of points with the same w_0^de. There are in total 46 slices of 816 points.

In the likelihood analysis, the copula likelihood (Eq. 6.30) is evaluated on a grid. The range of w_0^de is [-1.8, 0], with $\Delta w_0^\mathrm{de} = 0.04$. Concerning Ω_m and σ_8, only some particular values are chosen for evaluation in order to reduce the computing costs. This results in 816 points in the Ω_m-σ_8 plane, as displayed in Fig. 8.4, and the total number of parameter sets is 37536. For each parameter set, $N = 400$ realizations of our model are performed. Each realization produces data vectors for the Gaussian kernel, the starlet kernel, and MRLENS, so that the comparisons between cases are based on the same stochasticity. The aperture mass is not included here because of the time consuming convolution of the unbinned shear catalogue with the filter Q. A map example is displayed in Fig. 8.5 for the three cases and the input simulated κ field.

In the ABC analysis, all four filters are used. For the three first linear cases, the data vector $\boldsymbol{x}$ is composed of three scales. Each scale has the same nine S/N bins as in the likelihood analysis, which results in 27 bins in total. For MRLENS, $\boldsymbol{x}$ is a 6-bin κ histogram, also identical to the likelihood analysis (Table 8.3).

Concerning the ABC parameters, we use 1500 particles in the PMC process. The iteration stops when the success ratio of accept-reject processes falls below 1%. Finally, we test two distances. Between the sampled data vector $\boldsymbol{x}$ and the observed one, $\boldsymbol{x}^\mathrm{obs}$, we consider a
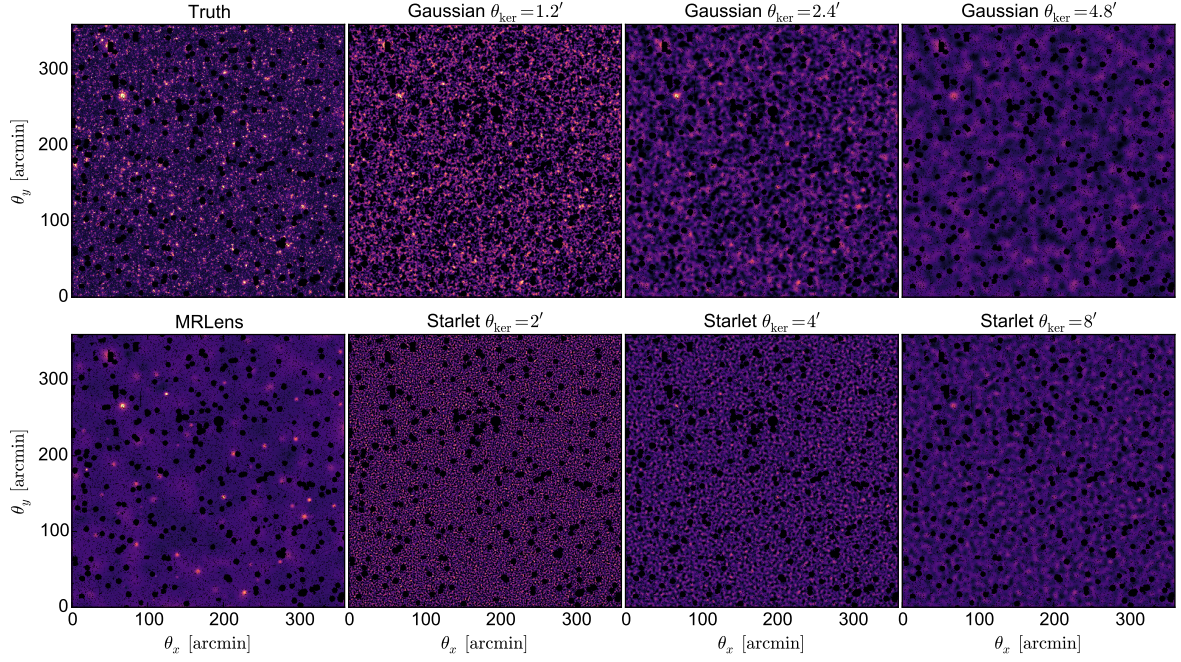
FIGURE 8.5: Maps taken from one of the simulations. The truth map is made by calculating κ_{proj} without noise. The panels of the rest are different filtering techniques applied on the map obtained from a KS inversion after calculating $\epsilon^{(\mathrm{o})} = (\epsilon^{(\mathrm{s})} + g_{\mathrm{proj}})/(1 + g_{\mathrm{proj}}^{*}\epsilon^{(\mathrm{s})})$. The black areas are masks. The unit of kernel sizes is arcmin.

Filter	θ_{ker} [arcmin] or α	Number of bins	d
Gaussian	$\theta_{\mathrm{ker}} = 1.2,\ 2.4,\ 4.8$	9 ν bins	27
Starlet	$\theta_{\mathrm{ker}} = 2,\ 4,\ 8$	9 ν bins	27
M_{ap} tanh	$\theta_{\mathrm{ker}} = 2.125,\ 4.25,\ 8.5$	9 ν bins	27
MRLens	$\alpha = 0.05$	6 κ bins	6

TABLE 8.3: Definition of the data vector $\boldsymbol{x}$ for PMC ABC runs. The 9 bins of ν are $[1, 1.5, 2, \ldots, 4, 4.5, 5, +\infty[$, and the 6 bins of κ are $[0.02, 0.03, 0.04, 0.06, 0.10, 0.16, +\infty[$. The symbol d is the total dimension of $\boldsymbol{x}$, and α stands for the input value of FDR for MRLens.

simplified distance D_1 and a fully correlated one D_2, which are respectively defined as

$$D_1\left(\boldsymbol{x}, \boldsymbol{y}\right) \equiv \sqrt{\sum_i \frac{\left(x_i - y_i\right)^2}{C_{ii}}}, \tag{8.15}$$

$$D_2\left(\boldsymbol{x}, \boldsymbol{y}\right) \equiv \sqrt{\left(\boldsymbol{x} - \boldsymbol{y}\right)^T \boldsymbol{C}^{-1} \left(\boldsymbol{x} - \boldsymbol{y}\right)}, \tag{8.16}$$

where C_{ii} and $\boldsymbol{C}^{-1}$ are now independent from cosmology, estimated under $(\Omega_{\mathrm{m}}, \sigma_8, w_0^{\mathrm{de}}) = (0.28, 0.82, -0.96)$.
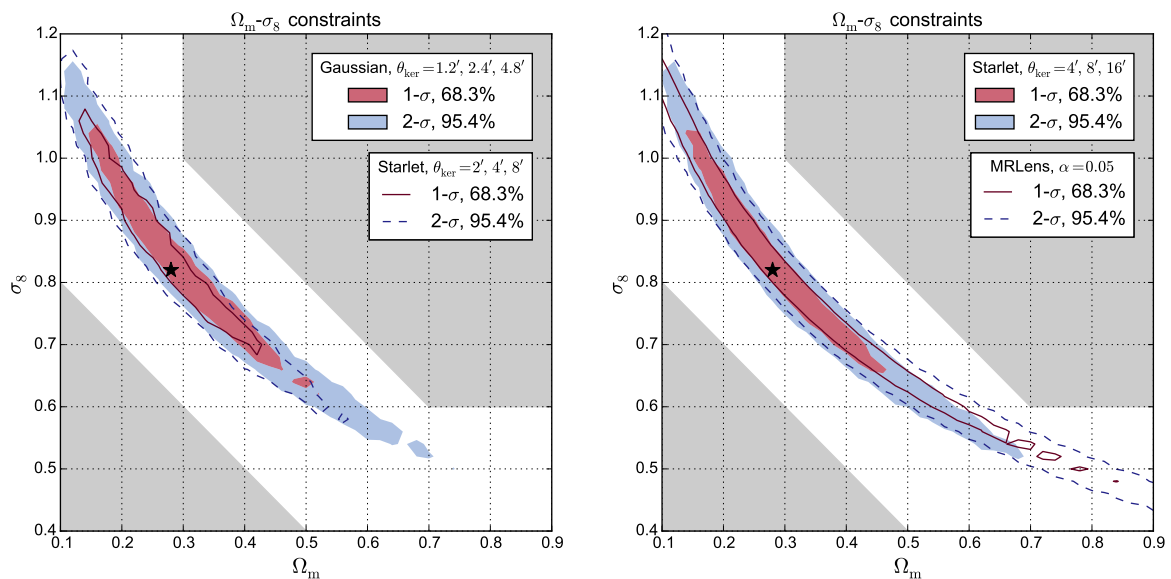
FIGURE 8.6: Ω_{m}-σ_8 constraints from four different cases. Left panel: the Gaussian case (colored regions) and the starlet case with three corresponding scales based on the noise level (solid and dashed contours, $\theta_{\mathrm{ker}} = 2$, 4, and 8 arcmin). Right panel: the starlet case with three corresponding scales based on number counts (colored regions, $\theta_{\mathrm{ker}} = 4$, 8, and 16 arcmin) and the MRLENS case (solid and dashed contours). The Gaussian and count-based starlet cases yield almost identical constraints. Between four cases, the best result is given by the noise-based starlet case. Black stars represent the input cosmology. Grey zones are excluded in this analysis.

8.5 Results

We again use the uncertainty on Σ_8 and the FoM on the Ω_{m}-σ_8 plane to qualify constraints. Here, Σ_8 is defined differently:

$$\Sigma_8 \equiv \left(\frac{\Omega_{\mathrm{m}} + \beta}{1 - \alpha}\right)^{1-\alpha} \left(\frac{\sigma_8}{\alpha}\right)^{\alpha}. \tag{8.17}$$

The motivation for this definition is to measure the "contour width" independently from α. With the common definition $\Sigma_8 \equiv \sigma_8 (\Omega_{\mathrm{m}}/\mathrm{pivot})^{\alpha}$, the uncertainty on Σ_8 under different α does not correspond to the same measure. The 1-σ error bar on Σ_8, $\Delta\Sigma_8$, is obtained using the same method as in Chap. 6.

8.5.1 With the copula likelihood

First, we test the maximum information that Gaussian kernels can extract. Table 8.4 shows the FoM from the marginalized likelihood. We can see that adding $\theta_{\mathrm{ker}} = 2.4$ and 4.8 arcmin to the filter with 1.2 arcmin has no significant effect on constraints. The constraints from the smallest filter are the most dominant ones among all.

Next, we use all three Gaussian scales as the reference for the comparisons with the starlet function. As mentioned in Sect. 8.4.1, for the Gaussian filter scales of 1.2, 2.4, and 4.8 arcmin, we chose scales for the starlet based on two criteria: for an equal noise level, these are 2, 4, and 8 arcmin, and for equal number counts the corresponding scales are 4, 8, and 16 arcmin. The results are shown in Fig. 8.6. For the equal-number-count criterion, we see that if each scale gives approximately the same number of peaks, the Ω_{m}-σ_8 constraints obtained from the Gaussian and the starlet are similar (colored regions in the left and right
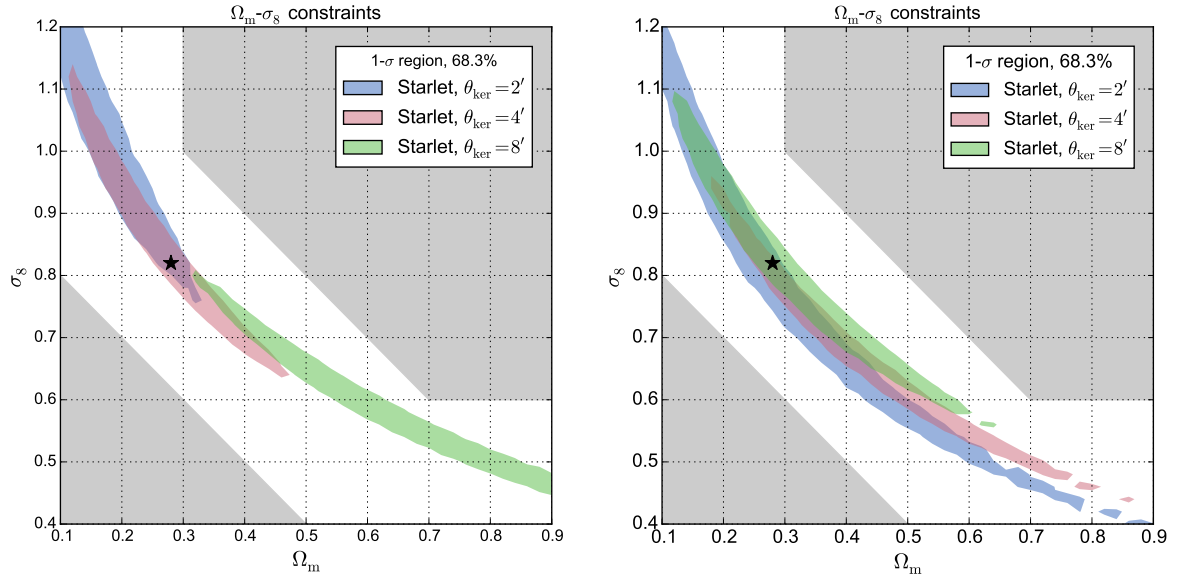
FIGURE 8.7: Ω_m-σ_8 1-σ region from individual scales of the starlet kernel. The plotted scales are 2 (blue), 4 (red), and 8 arcmin (green). The left panel is the result from the observation vector used in the analysis; the right panel represents the constraints from another realization of the observation. We see that the behaviors between different scales is more likely due to stochasticity. Grey zones are excluded in this analysis.

panels). However, the starlet kernel leads to tighter constraints than the Gaussian when we match the same noise levels (lines and colored regions in the left panel). This results suggests that compensated kernels could be more powerful to extract cosmological information than non-compensated filters.

We also draw constraints from individual scales of the starlet filter in (Fig. 8.7, left panel). It shows a very different behavior and seems to suggest that different scales could be sensitive to different cosmologies. However, this is actually a stochastic effect. We verify this statement by redoing the constraints with other observation vectors. It turns out that the scale-dependent tendency disappears (Fig. 8.7, right panel). Nevertheless, when different cosmologies are preferred by different scales, the effect is less pronounced for the Gaussian filter. This is likely due to the fact that the starlet is a compensated filter, which is a band-pass function in the Fourier space. Since different filtering scales could be sensitive to different mass ranges of the mass function, band-pass filters weaken the correlations between scales (Fig. 8.9) and separate better the multiscale information. The stochasticity of the observation vector suggests that the simulated field of view is rather small. While this should not affect the filter comparison nor contour sizes, actual cosmological constraints seem to require substantially larger data sets.

The right panel of Fig. 8.6 shows the constraints from nonlinear filtering using MRLENS (solid and dashed lines). We observe that MRLENS conserves a strong degeneracy between Ω_m and σ_8. The reasons for this result are various. First, large and small scales tend to be sensitive to different halo masses which could help break the degeneracy. Using the combined strategy loses this advantage. Second, a strict FDR has been chosen. This rules out most of the spurious peaks, but also a lot of the signal. Third, as mentioned before, it is inappropriate to define signal-to-noise ratio when the filter is not linear. As a consequence, it is hardly possible to find bins for κ peaks which are equivalent to ν bins in linear filtering. This is supported by Fig. 8.5, where we observe less peaks in the MRLENS map than in the other

Filter	θ_{ker} [arcmin] or α	$\Delta\Sigma_8$	FoM
Gaussian	$\theta_{\mathrm{ker}} = 1.2$	0.045	19.1
Gaussian	$\theta_{\mathrm{ker}} = 1.2, 2.4, 4.8$	0.046	20.7
Starlet	$\theta_{\mathrm{ker}} = 2, 4, 8$	0.046	23.4
Starlet	$\theta_{\mathrm{ker}} = 4, 8, 16$	0.044	21.2
Starlet	$\theta_{\mathrm{ker}} = 2, 4, 8, 12, 16$	0.045	24.8
MRLens	$\alpha = 0.05$	0.046	16.2

TABLE 8.4: Quality indicators for Ω_{m}-σ_8 constraints with likelihood. All cases figured below use number counts on g peaks. The quantity $\Delta\Sigma_8$ stands for the width of the contour, while the FoM is related to the area. In our study, combining five scales of starlet yield the best result in terms of FoM.
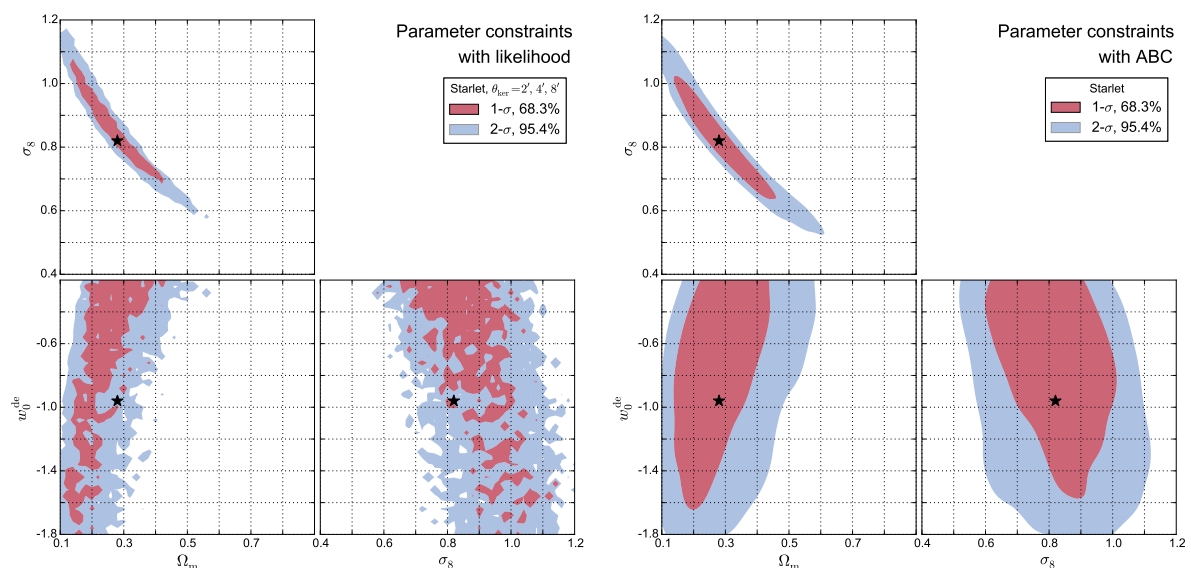

FIGURE 8.8: Ω_{m}-σ_8-w_0^{de} constraints using starlet with three scales. Left panel: constraints with the likelihood. Fluctuations on both lower panels are due to usage of the copula likelihood. Right panel: constraints with ABC using the distance D_2. Each panel represents the marginalization over one of the three parameters. Black stars are the input cosmology. As far as w_0^{de} is concerned, the constraints are weak, but the degeneracies are clear.

maps. Last, because of a low number of peaks, the binwidths need to be enlarged to contain larger number counts and to get closer to a Gaussian distribution, and large binwidths also weaken the signal.

A possible solution for exploring the MRLens technique is to enhance the FDR and to redesign the binning. By increasing the number of peaks, thinner bins would be allowed. Another solution to better account for rare events in the current configuration is to use the Poisson likelihood. Finally, one could adopt the separated strategy, i.e. turning back to the methodology used by Pires et al. (2009a) that consists in estimating the peak abundance in the different scales before final reconstruction. In this study, the comparison between "linear and nonlinear techniques" is basically the one between the "separated and combined strategies".

Table 8.4 measures numerical qualities for constraints with different filtering techniques. It indicates that the width of contours does not vary significantly. The tightest constraint that we obtain is derived from a compensated filter.

Filter	Constraints	$\Delta\Sigma_8$	FoM
Gaussian	Likelihood	0.046	20.7
Gaussian	ABC, D_1	0.043	16.3
Gaussian	ABC, D_2	0.059	11.7
Starlet	Likelihood	0.054	23.4
Starlet	ABC, D_1	0.050	15.5
Starlet	ABC, D_2	0.054	15.7
M_{ap} tanh	ABC, D_1	0.037	19.4
M_{ap} tanh	ABC, D_2	0.043	15.5
MRLens	Likelihood	0.046	16.2
MRLens	ABC, D_1	0.045	11.5
MRLens	ABC, D_2	0.045	12.5

TABLE 8.5: Quality indicators for Ω_{m}-σ_8 constraints with PMC ABC. The quantity $\Delta\Sigma_8$ stands for the width of the contour, while the FoM is related to the area. ABC is used with two different distances D_1 and D_2 respectively given by Eqs. (8.15) and (8.16). Here, we also put values from likelihood constraints using the same scales in this table for comparison. The kernel sizes for linear methods are defined in Table 8.3.

Regarding results for w_0^{de}, we show a representative case of starlet with $\theta_{\mathrm{ker}} = 2$, 4, and 8 arcmin. The left panel of Fig. 8.8 presents the marginalized constraints of each doublet of parameters that we study. Those containing w_0^{de} are noisy because of the usage of the copula likelihood. We may see that the current configuration of our model does not allow to impose constraints on w_0^{de}. To measure this parameter, it could be useful to perform a tomography analysis to separate information of different stages of the late-time Universe. Nevertheless, our results successfully highlight the degeneracies of w_0^{de} with two other parameters. We fit the posterior density with:

$$I_1 = \Omega_{\mathrm{m}} - a_1 w_0^{\mathrm{de}}, \tag{8.18}$$

$$I_2 = \sigma_8 + a_2 w_0^{\mathrm{de}}. \tag{8.19}$$

We obtain for the slopes $a_1 = 0.108$ and $a_2 = 0.128$ for the left panel of Fig. 8.8. The results for the other filter functions are similar.

8.5.2 With PMC ABC

We perform parameter constraints using the PMC ABC algorithm for our four cases. In the right panel of Fig. 8.8, we show the results derived from the starlet case using the fully correlated distance D_2. The contours are marginalized posteriors for all three pairs of parameters. They show the same degeneracy as we have found with the likelihood. We measure a_1 and a_2 from the ABC posteriors and obtain $a_1 = 0.083$ and $a_2 = 0.084$.

Using the same starlet filters, we compare two distances used for PMC ABC runs. When D_1 is used with the starlet, i.e. data are treated as if uncorrelated, we find that the contour sizes do not change (see Table 8.5) compared to D_2. For the Gaussian case, however, constraints from D_1 are tighter than those from D_2. This phenomenon is due to the off-diagonal elements of the covariance matrix. For non-compensated filters, the cross-correlations between bins are much stronger, as shown in Fig. 8.9. If these cross-correlations are ignored, the repeated peak counts in different bins are not properly accounted for. This overestimates the
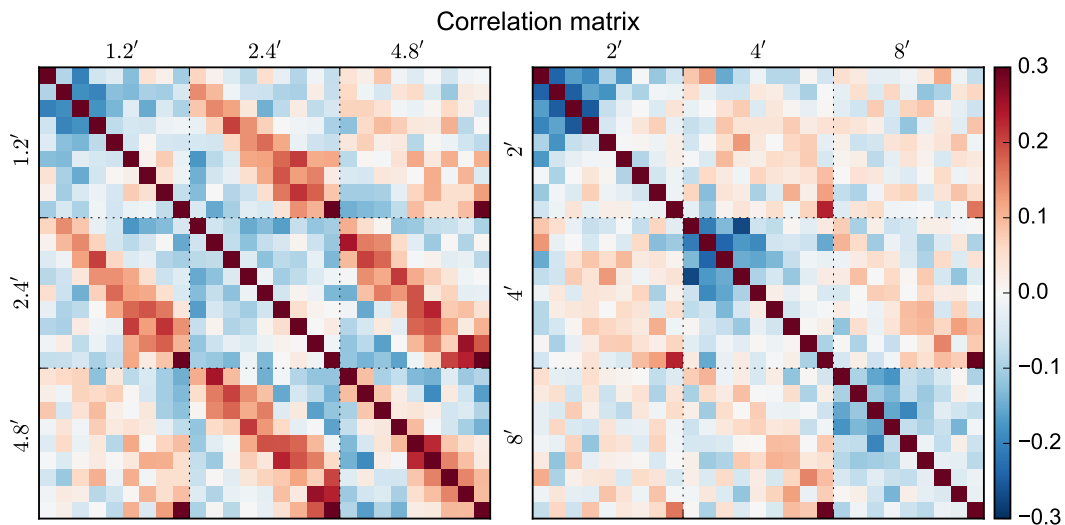
FIGURE 8.9: Correlation coefficient matrices under the input cosmology. Left panel: the Gaussian case with $\theta_{\mathrm{ker}} = 1.2$, 2.4, and 4.8 arcmin. Right panel: the starlet case with $\theta_{\mathrm{ker}} = 2$, 4, and 8 arcmin. Each of the 3×3 blocks corresponds to the correlations between two filter scales. With each block, the S/N bins are [1, 1.5, 2, ..., 5, $+\infty$[. The data vector by starlet is less correlated.

additional sensitivity to massive structures, and therefore produces overly tight constraints. As shown in Fig. 8.9, in the Gaussian case, adjacent filter scales show a 20–30% correlation. The blurring of the off-diagonal stripes indicate a leakage to neighboring S/N bins due to noise, and the fact that clusters produce WL peaks with different S/N for different scales. On the contrary, in the case of the starlet, except for the highest S/N bin there are negligible correlations between different scales.

Table 8.5 shows the ABC constraints from both the aperture mass and the starlet. We find that the FoM are close. However, in Fig. 8.10, we see that the contours from the aperture mass is shifted toward high-Ω_{m} regions. The explanation for this shift is once again the stochasticity. We simulated another observation data vector for M_{ap}, and the maximum-likelihood point for different methods do not coincide.

From Table 8.5, one can see that the difference between MRLENS and linear filters using ABC is similar to using the likelihood. This suggests once again that the combined strategy leads to less tight constraints than the separated strategy. Note that we also try to adjust α and run PMC ABC. However, without modifying the κ bin choice, the resulting constraints do not differ substantially from $\alpha = 0.05$.

Finally, we show the likelihood and ABC constraint contours for the Gaussian and starlet cases in Fig. 8.11. It turns out that ABC contours are systematically larger in the high-Ω_{m}, low-σ_8 region. This phenomenon was not observed previously. We speculate that by including a third parameter w_0^{de} the contour becomes less precise, and ABC might be more sensitive to this effect. Note also that KDE is a biased estimator of posteriors. It smooths the posterior and makes contours broader. Nevertheless, the ABC and likelihood constraints agree with each other. To be free from the bias, a possible alternative is to map the samples to a Gaussian distribution via some nonlinear mapping techniques (Schuhmann et al. 2016).
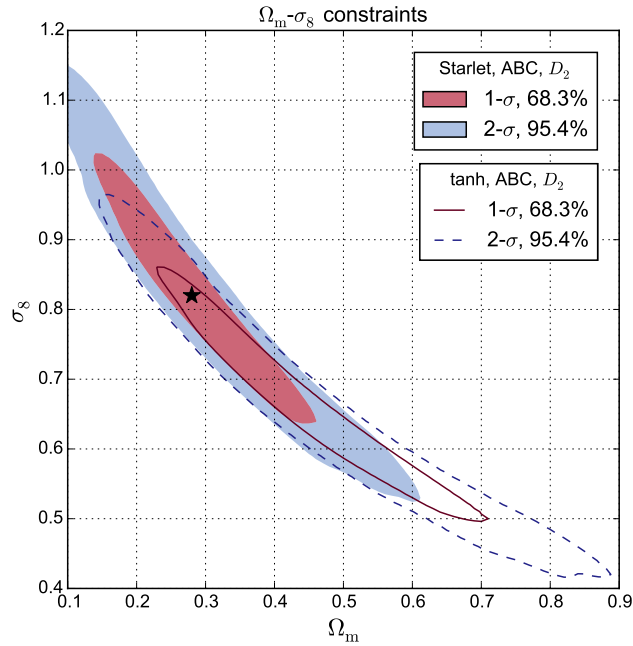
FIGURE 8.10: ABC Ω_{m}-σ_8 constraints from the starlet and the aperture mass. The distance D_2 is used in both cases. The black star is the input cosmology. The difference between two cases is that another observation data vector is created for the aperture mass and the direct comparison is not valid anymore.

Summary

In this chapter, a new standard of comparisons between various filtering techniques has been presented. Comparing directly the constraint contours instead of purity and completeness of the cluster detection leads to a more direct measurement of cosmological information extraction.

The study shows a preference of compensated filters rather than non-compensated ones. Also, the separated strategy outperforms the combined strategy in terms of multiscale information extraction.

Constraints from the likelihood seem to be tighter than from ABC. Since in Sect. 6.4 we found that the copula likelihood closely approximates the true one, ABC probably overestimate the true parameter errors.

Without tomography, w_0^{de} could not be constrained tightly. However, we observe that it is degenerate with Ω_{m} and σ_8.

Our model has been improved to be adapted to more realistic observational conditions. We will now present a data application in the next chapter.
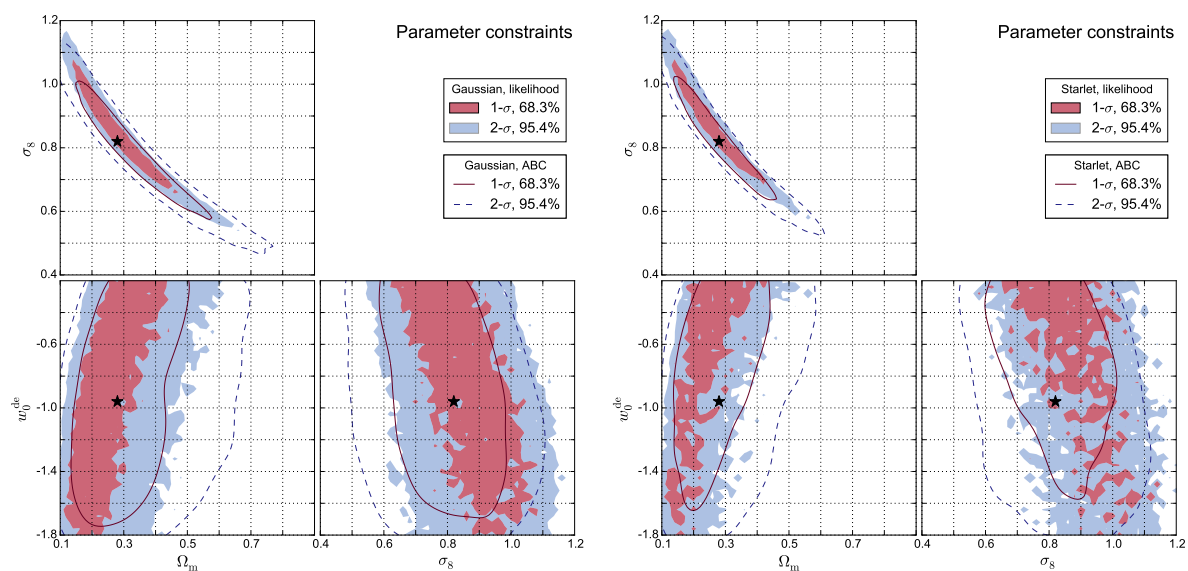
FIGURE 8.11: Comparison of Ω_m-σ_8-w_0^de constraints between likelihood and ABC. Left panel: constraints with Gaussian smoothing. Right panel: constraints with starlet filtering. Although ABC tolerates higher Ω_m and lower σ_8 in both cases, two methods agree with each other.

Chapter IX

Data applications

Overview

In this chapter, we apply our analysis to three data sets. The cosmological constraints are performed using our peak-count model and ABC. We will show preliminary results, possible improvements, and the perspectives of this work at the end.

9.1 Data descriptions

Three data sets are processed: the full data from the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS), the data releases 1 and 2 from the Kilo-Degree Survey (KiDS DR1/2), and the science verification data from the Dark Energy Survey (DES SV).

CFHTLenS

The CFHTLenS data (Heymans et al. 2012; Erben et al. 2013) are composed of four separated fields W1, W2, W3, and W4. The total area is 154 deg^2. The total unmasked area is about 126 deg^2. Shape measurement was done with LENSFIT, which is a model-fitting algorithm (Miller et al. 2013). Galaxies are selected with redshift, measurement weight, and mask flags, respectively as $0.2 < $ `Z_B` $ < 1.3$, `weight` > 0, and `MASK` < 2 [♯1]. This results in 6.1 million galaxies. The raw density is 13.53 arcmin^{-2}.

To process peak counts, the W1 field, which is relatively large, is cut into four patches. Also, the W4 field which has a irregular shape is separated into two patches. Thus, the whole data set is represented by eight patches of different sizes (Fig. 9.1). For each patch, border effects are carefully removed. This means that some galaxies near the edges are sacrificed and only peaks in the inner area are counted. We define it as the *effective area* in the same way as in previous chapters, which is 112 deg^2 in this case.

[♯1] All science analyses of the CFHTLenS team are performed with `MASK` < 2. These objects can be safely used for most scientific purposes. (Erben et al. 2013)
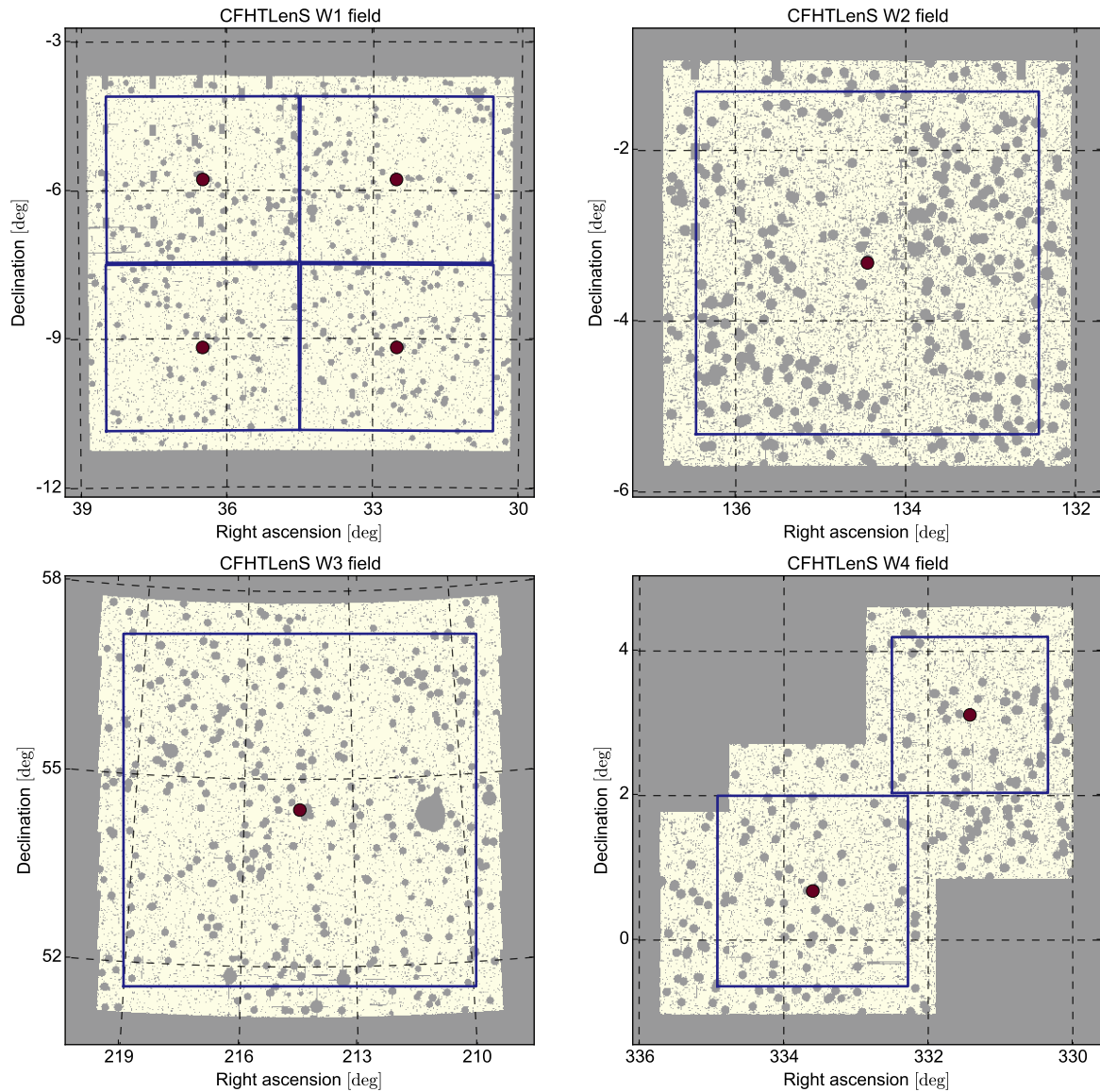
FIGURE 9.1: Illustrations of all eight patches (blue frames) from CFHTLenS. The four panels represent respectively the W1, W2, W3, and W4 fields. Red points are the center of patches. Unmasked areas are showed in yellow.

KiDS DR1/2

KiDS DR1/DR2 (de Jong et al. 2015) contains 109 tiles of 1 deg^2 all near the equator. Concerning the lensing catalogue, KiDS uses the same algorithm as CFHTLenS. We select sources following the criteria suggested by Kuijken et al. (2015): `MAN_MASK` $= 0$, $0.005 <$ `Z_B` < 1.2, `weight` > 0, and `SNratio` > 0. In addition to these, due to calibration limits, we include also `c1_best` > -50 and `c2_best` > -50 to cut out values of -99 (Kuijken et al. 2015). The final selection includes 2.4 million galaxies distributed on an unmasked area of 75 deg^2. The raw density is 8.87 arcmin^{-2}.

Regarding the fact that (1) the KiDS data are more contaminated by masks and (2) some 1-deg^2 tiles are missing from a contiguous region, we only retain a small part of the whole data set. This is organized as four patches with a total area of 41 deg^2 (Fig. 9.2). After
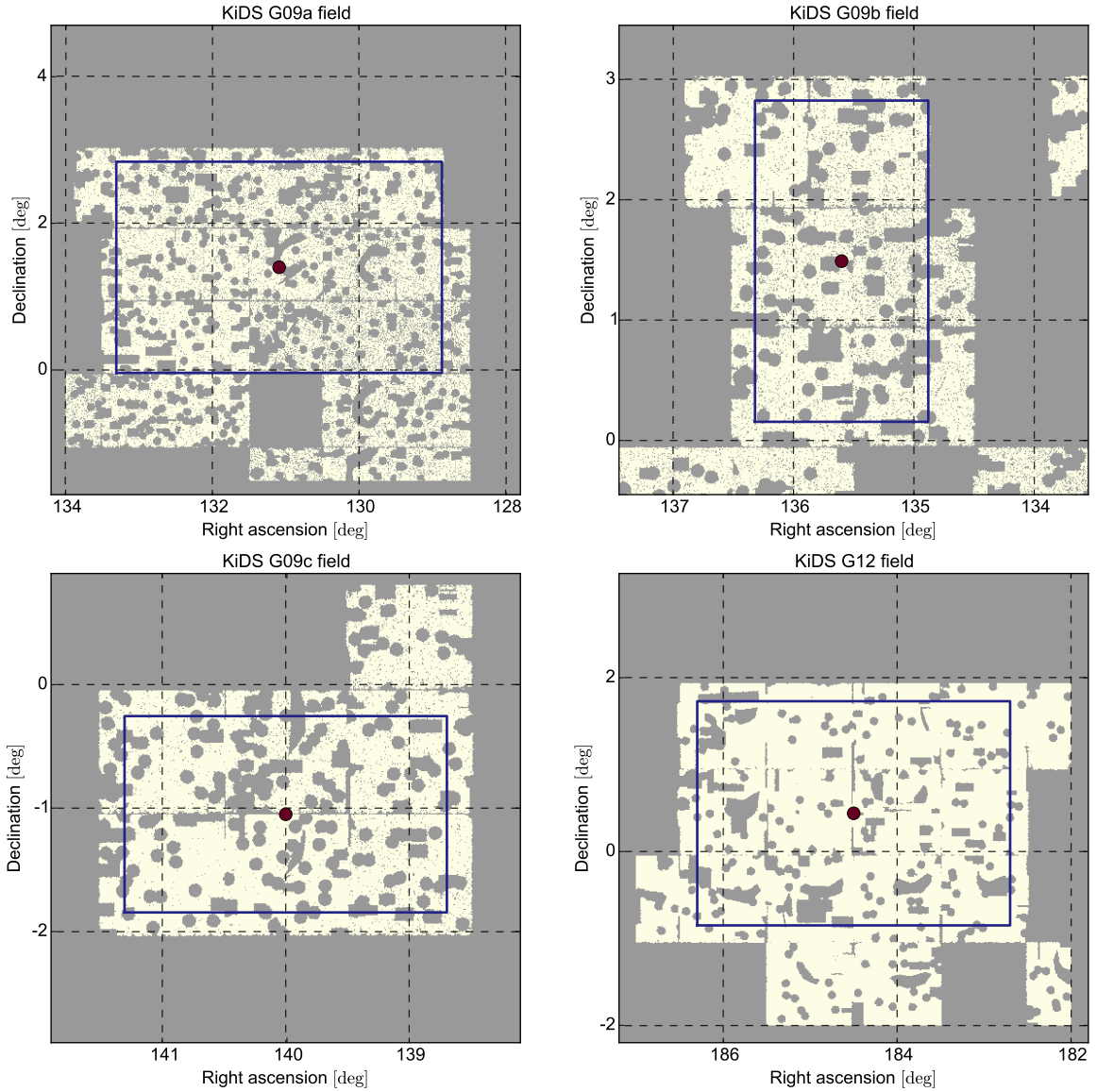
FIGURE 9.2: Illustrations of all four patches (blue frames) from KiDS DR1/2. Three of them are taken from the G09 field. The rest is taken from the G12 field. Red points are the center of patches. Unmasked areas are showed in yellow.

cutting out edges, the effective area is 30 deg^2.

DES SV

The DES SV data (Dark Energy Survey Collaboration et al. 2016; Jarvis et al. 2016) contains a large field named SPT-E and some other small fields. Here we only focus on the SPT-E field, which covers an unmasked area of 138 deg^2. We use the NGMIX catalogue provided by the collaboration. The flag NGMIX_FLAG $= 0$, which is a combination of several flags, validates that a source has a good NGMIX measurement [#2]. An additional criterion is the redshift information, with $0.3 <$ MEAN_PHOTOZ < 1.3, suggested by Kacprzak et al. (2016). The total

---

[#2]See http://des.ncsa.illinois.edu/.
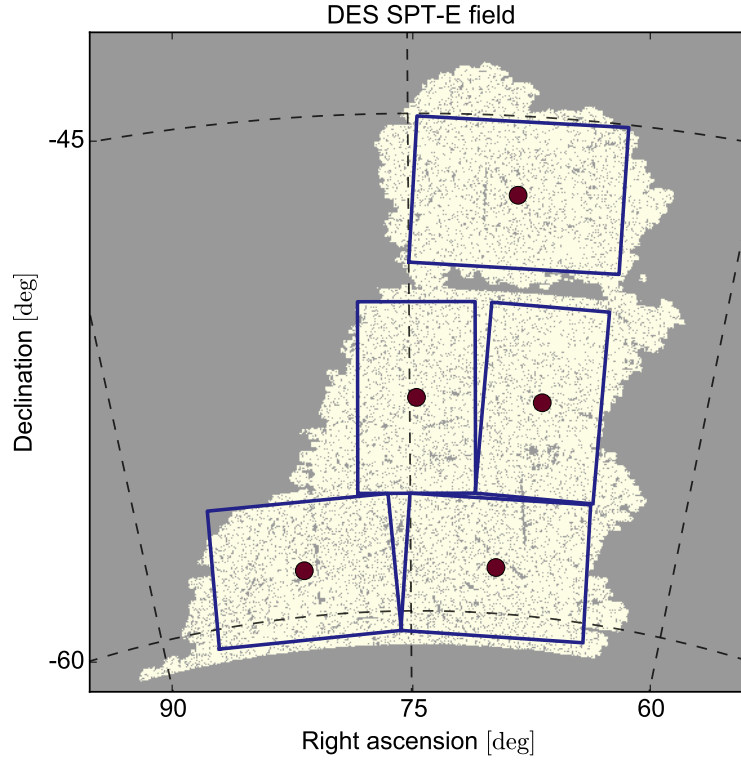
FIGURE 9.3: Illustrations of all five patches (blue frames) from DES SV. Red points are the center of patches. Unmasked areas are showed in yellow.

number of galaxies is 3.3 millions, with raw density 6.63 arcmin^2.

In this study, the SPT-E field is divided into five patches (Fig. 9.3). The resulting effective area is 115 deg^2.

9.2 Methodology

The analysis is processed with the same method as described in Sect. 8.4.2, with $M_{\min} = 2 \times 10^{12}\ M_\odot$. Each survey is modelled separately with the corresponding parameters such as n_{gal}, σ_ϵ, θ_{pix}, filtering scales, and the source redshift distribution (see Table 9.1). At the end, three peak histograms are joined. The details are described in the following paragraphs.

Source redshift distribution

The redshift distribution of the sources is assumed to be a general gamma distribution:

$$p(z) \propto \left(\frac{z}{z_0}\right)^{\alpha_{\mathrm{gal}}} \exp\left(-\left(\frac{z}{z_0}\right)^{\beta_{\mathrm{gal}}}\right), \tag{9.1}$$

where the normalization is computed for the interval $[z_{\min}, z_{\max}]$. For each survey, $z_{\min}$ and $z_{\max}$ are given by the suggested flags. Then, we fit α_{gal}, β_{gal}, and z_0 to the redshift distribution of the catalogues.

Parameter	Symbol	CFHTLenS	KiDS DR1/2	DES SV
Source distribution parameter	$z_{\min}$	0.2	0.005	0.3
Source distribution parameter	$z_{\max}$	1.3	1.2	1.3
Source distribution parameter	$\alpha_{\rm gal}$	0.474	1.255	2.958
Source distribution parameter	$\beta_{\rm gal}$	5.990	0.698	1.048
Source distribution parameter	z_0	1.157	0.128	0.217
Galaxy number density [arcmin^{-2}]	$n_{\rm gal}$	13.53	8.87	6.63
Ellipticity dispersion	σ_ϵ	0.449	0.516	0.400
Pixel size [arcmin]	$\theta_{\rm pix}$	0.72	0.9	1.05
Number of filtering scales	-	4	3	3

TABLE 9.1: List of parameter values adopted in the study of this chapter.

Ellipticity dispersion and galaxy density

The noise from the catalogue has two origins: intrinsic shape dispersion and shape measurement errors. Unfortunately, it is impossible to separate these two components from observations. For this reason, cosmologists define the effective number density, such that when this quantity is combined with the assumed intrinsic dispersion, the total variance will match the one in the catalogue. Therefore, the effective density is always smaller than the raw density. The less the galaxies, the higher the Poisson noise; and this Poisson part will account for the measurement errors neglected in the effective formalism.

However, this is actually a redundant step. First, the true shape dispersion is not Gaussian, neither are the measurement errors. Separating both components in the Gaussian formalism seems very unrealistic. Second, the intrinsic dispersion is never known. Although Jarvis et al. (2016) claim that it is possible via a fine-tuned combination of weights, the results obtained by this technique are not consistent between different surveys. Finally, we never need the value of shape dispersion alone in the model. Modelling the combination of both components as a Gaussian noise is much simpler than considering them with separate models and assuming further approximations.

As a result, we adopt a different approach to determine the density and the dispersion. The idea is simple: given a catalogue with its total variance, we look for the Gaussian dispersion which corresponds to the raw density, so that the total variance matches. This leads to two advantages. First, only one assumption is needed: the Gaussianity of the combination of the intrinsic dispersion and the shape measurement. Second, the number density that will be used in our model is larger. This reduces the bias generated by the KS inversion when the distribution of sources is too irregular.

More explicitly, the total variance is defined as the empirical weighted variance taken the multiplicative bias into account:

$$\sigma_{\rm tot}^2 = \frac{\sum_i w_i^2 \left(\epsilon_{i,1}^{(c)2} + \epsilon_{i,2}^{(c)2} \right)}{\sum_i w_i^2 (1 + m_i)^2}, \tag{9.2}$$

where $\epsilon_{i,j}^{(c)} \equiv \epsilon_{i,j}^{(o)} - c_{i,j}$ is the j-th component of the observed ellipticity $\epsilon_i^{(o)}$ of the i-th galaxy corrected with the additive bias c_i, w_i is the weight, m_i is the multiplicative bias, and the sum runs over all galaxies in the catalogue. The Gaussian dispersion σ_ϵ corresponding to the

Data set	CFHTLenS
Filter scales θ_{ker} [arcmin]	1.44, 2.88, 5.76, 11.52
S/N bins	[2.5, 2.75, 3.0, 3.25, 3.5, 3.75, 4.0, 4.25, 4.75, $+\infty$[
Data set	KiDS DR1/2
Filter scales θ_{ker} [arcmin]	1.8, 3.6, 7.2
S/N bins	[2.5, 2.75, 3.0, 3.25, 3.5, $+\infty$[
Data set	DES SV
Filter scales θ_{ker} [arcmin]	2.1, 4.2, 8.4
S/N bins	[2.5, 2.75, 3.0, 3.25, 3.5, 3.75, 4.0, 4.5, $+\infty$[

TABLE 9.2: List of scales and bins used in the study of this chapter. The dimension of the data vector is $4 \times 9 + 3 \times 5 + 3 \times 8 = 75$.

raw density satisfies

$$\sigma_{\mathrm{tot}}^2 = \frac{\sigma_\epsilon^2}{2N_{\mathrm{gal}}}, \tag{9.3}$$

where N_{gal} is the total number of sources. In this case, the derived σ_ϵ can have a relatively large number. In practice, we truncate at $|\epsilon| = \pm 1$ to avoid unphysical values. The violation of the Gaussian variance by the truncation is ignored.

Pixel size and filter scales

The pixel size θ_{pix} is determined based on n_{gal}. The goal is to avoid having pixels with a low number of galaxies. We expect θ_{pix} to satisfy

$$P_{\bar{N}}(N = 0) + P_{\bar{N}}(N = 1) + P_{\bar{N}}(N = 2) \leq 0.03, \tag{9.4}$$

where $\bar{N} = n_{\mathrm{gal}}\theta_{\mathrm{pix}}^2$ is the expected number of galaxies in a pixel, and $P_\lambda(N)$ is the probability associated at number N for the Poisson distribution with parameter λ. This means that we expect the probability of having less than three galaxies in a pixel to be smaller than 3%. The values in Table 9.1 satisfy this criterion.

Concerning the filter, we only use the starlet filtering in this study. The filter scales are chosen to be some even multiples of the pixel size. The reason is that if $\theta_{\mathrm{ker}} = 2N\theta_{\mathrm{pix}}$ where $N = 1, 2, \ldots$, then the starlet is strictly compensated in its discrete sum. These values are some natural choices for wavelet functions. We pick four scales for CFHTLenS and three scales for the others. This results in 1.44, 2.88, 5.76, and 11.52 arcmin for CFHTLenS; 1.8, 3.6, and 7.2 arcmin for KiDS; and 2.1, 4.2, and 8.4 arcmin for DES.

Data vector

The data vector is defined as the concatenation of peak histograms of all scales from three surveys. The binwidth is defined slightly differently depending on the survey (see Table 9.2). The reason for this is to have enough number counts in each bin ($\gtrsim 8$) such that the Poisson distribution can be approximated by a Gaussian one. This results in a data vector of dimension 75. The correlation matrix of this data vector is shown in Fig. 9.4.
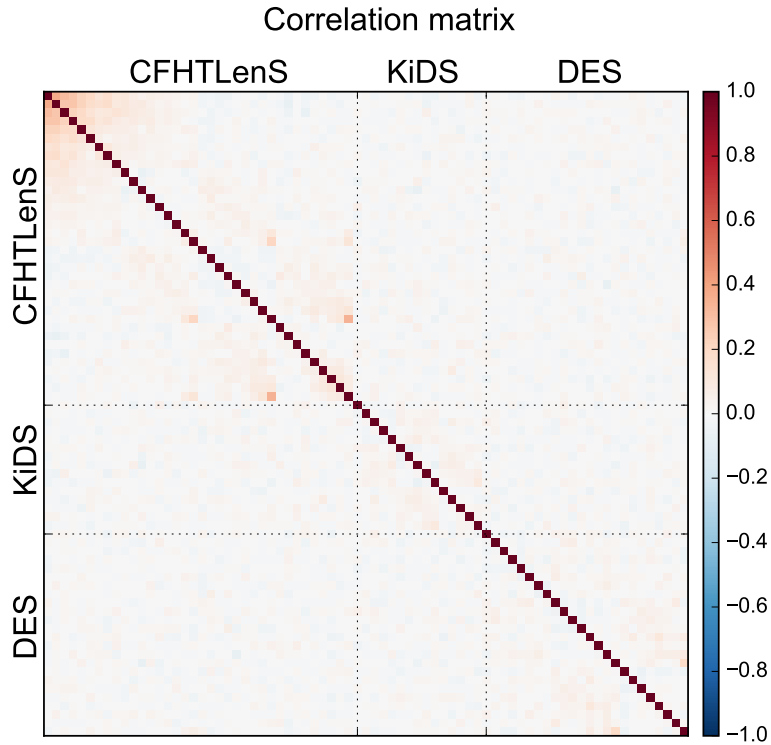
FIGURE 9.4: Correlation matrix from the data vector defined in this study. The components are very weakly correlated with each other.

Parameter sampling

We only perform the ABC analysis in this chapter. Two runs have been carried out. The first focuses on three free parameters, $\Omega_{\rm m}$, σ_8, and $w_0^{\rm de}$. The second one includes two additional parameters, c_0 and β, which define the halo *M-c* relation (Eq. 3.93).

Concerning the configuration of ABC, we take a flat prior in the studied parameter space. The number of particles is 2400. The shutoff parameter is set to 1%. The distance is the same as Eq. (8.16), accounting for the full correlation.

9.3   Preliminary results

Figure 9.5 shows the primary results that we obtain from the joint data constraints. In this figure, we only focus on $\Omega_{\rm m}$ and σ_8 from the run with three free parameters. As ϵ decreases, the contour size reduces. The constraints from the last iteration are shown in Fig. 9.6. The orange curve shows the best power-law fit (i.e. best Σ_8 fit, Eq. 8.17) between $\Omega_{\rm m}$ and σ_8. It indicates a very good agreement with the Planck cosmology ($\Omega_{\rm m} = 0.308\pm0.012$, $\sigma_8 = 0.8149\pm0.0093$). The contour is consistent with constraints from other lensing surveys using two-point statistics (CFHTLenS: Kilbinger et al. 2013, Joudaki et al. 2016; KiDS: Hildebrandt et al. 2016; DES: The Dark Energy Survey Collaboration et al. 2015) or peaks (CFHTLenS: Liu J. et al. 2015a; CFHT Stripe 82: Liu X. et al. 2015b; DES: Kacprzak et al. 2016). However, since our results are only preliminary, the comparisons are not shown. Here, we only give the characteristics of the preliminary constraints: the error on the orange curve is $\Delta\Sigma_8 = 0.13$, obtained by the
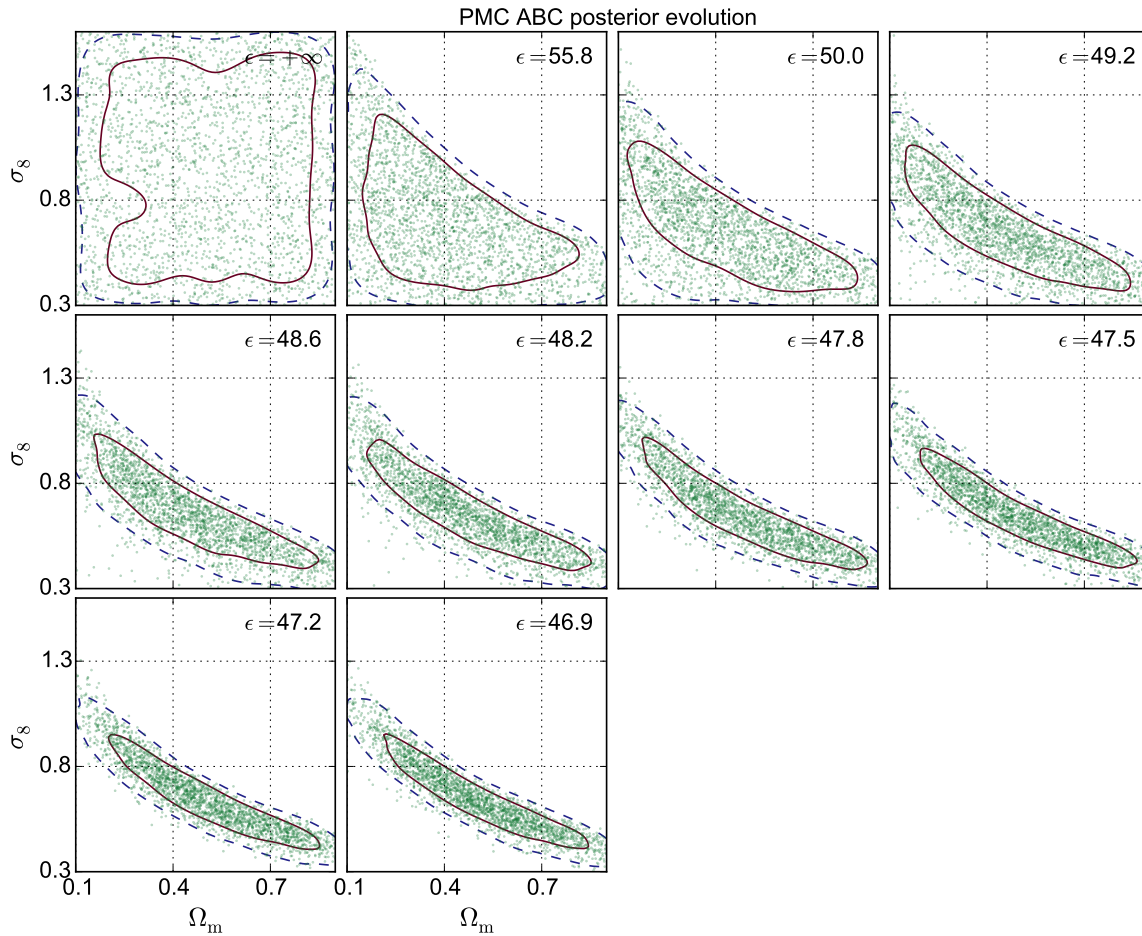
FIGURE 9.5: Preliminary result of Ω_{m}-σ_8 constraints with CFHTLenS-KiDS-DES joint data sets. This figure shows the evolution of the PMC ABC. The blue solid and the red dashed lines represent respective 1- and 2-σ contours. Green points are accepted ABC particles.

same method as Chap. 8, and the FoM is 5.2.

In this preliminary result, the constraints seem to be larger than expected. If we compare the FoM from data with those from Table 8.5, then the difference is about a factor of 3. Since the settings for the results from Table 8.5 have been designed to be very similar to the CFHTLenS data set, we believe that they can be compared directly. The reason of these looser contours is believed to be the selection of S/N bins and the bias from our model. A quick comparison (not plotted) with the Mice simulations (5200 deg^2) indicates that the peak function given by our model is tilted from the one derived from N-body simulations. Typically, our model overestimates high peaks and underestimates low peaks. If this is the case, then changing the input cosmology of our model, which leads to a global change in all bins of the peak functions, could result in similar differences with regard to the observation, therefore degrades the constraint power. In this scenario, the bad accuracy reduces the precision, and this is susceptible to be the case of our preliminary result. For the future, the improvement will be to compare closely our model with the Mice simulations to quantify the bias, and to adopt only bins with weak bias for constraints. With this refinement, we expect that the final results yield a similar FoM and $\Delta\Sigma_8$ to the values from Table 8.5.

Figure 9.7 shows all marginalized constraints from the run with three free parameters.
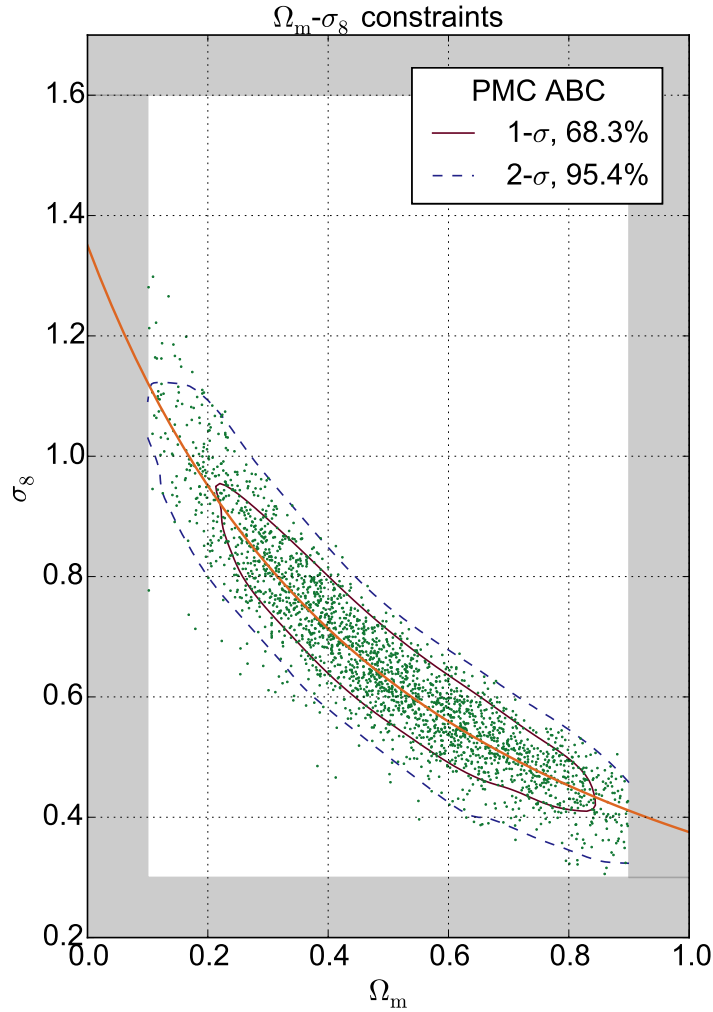
FIGURE 9.6: Preliminary result of Ω_m-σ_8 constraints with CFHTLenS-KiDS-DES joint data sets. This figure shows the posterior from the final iteration. The blue solid and the red dashed lines represent respective 1- and 2-σ contours. Green points are accepted ABC particles. The orange curve is the best power-law fit.

The degeneracy between w_0^{de} and other parameters is less significant because of ABC. Kernel density estimation smooths the contours. In Fig. 9.8, we show the constraints with five free parameters. We notice that the constraint on c_0 is very broad. However, concerning β, the values below -0.6 are already been excluded by the preliminary study. By including two additional parameters, the constraints on Ω_m and σ_8 become larger, which is expected.

9.4 Perspectives

In spite of the three-year investment, several questions concerning peak-count modelling still remained unanswered. Below is a list non-exhausted of ideas and perspectives that future studies could investigate.

Halo modelling   All along the thesis, we adopt the spherical NFW modelling for halo density profiles. However, N-body simulations indicate well the asphericity of dark-matter halos.

FIGURE 9.7: Preliminary result of Ω_m-σ_8-w_0^de constraints with CFHTLenS-KiDS-DES joint data sets.
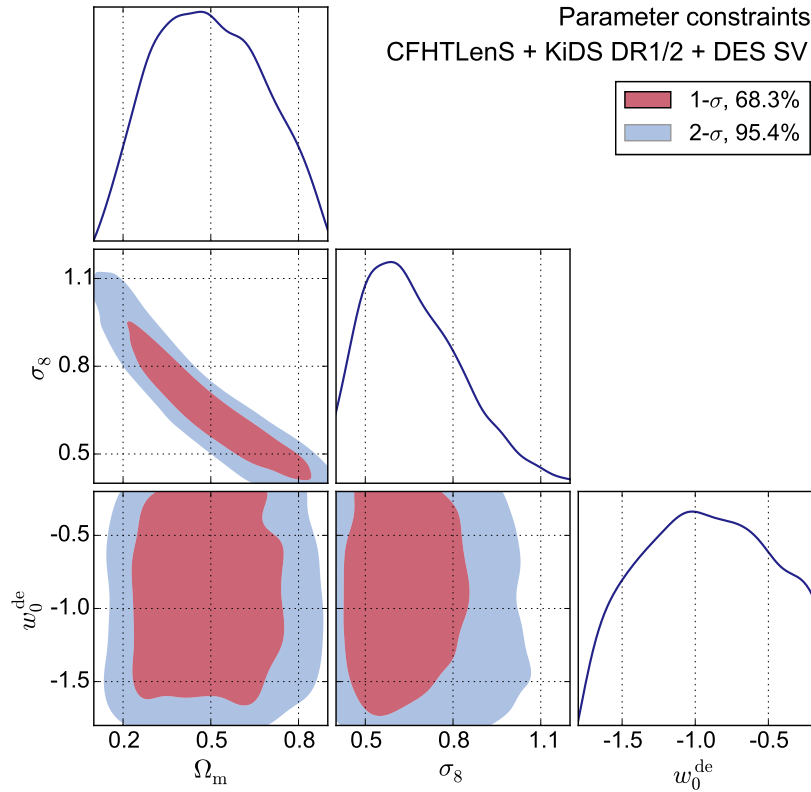
Without reproducing their true shapes, triaxial halos or substructures can be a more realistic option for lens modelling. Meanwhile, baryonic effects can also modify the profiles. These different possible improvements concerning halo geometry are worth taking a closer look. Next, in Fig. 5.5, we have already seen that the M-c relation of halos suffers from a huge scatter. Therefore, taking this intrinsic scatter into account can lead to a more accurate modelling than a deterministic relation. Finally, halo correlation is another subject to investigate in detail. We have seen that neglecting correlation leads to a decrease of peak counts which is compensated by the replacement of N-body halo profiles with their NFW proxies. Why do these two effects vary in opposite ways? Do they really compensate? These are some questions to investigate for the future.

Source modelling The improvement for source modelling includes at least three aspects: photo-z, galaxy clustering, and intrinsic alignment. The actual photometric estimation of galaxy redshifts is scattered and can be biased. For real surveys, we need to correct the bias for peak counts, and to take the propagation of the scatter into account. Meanwhile, instead of using randomized sources as in this work, adopting the exact source catalogue would reduce the stochasticity. Using the survey source catalogue will also account for galaxy clustering, which is a potential source of bias for peak counts (Kacprzak et al. 2016). Finally, intrinsic alignment plays an important role in weak lensing analyses. Understanding its impacts on weak-lensing peaks will be crucial for future large and high-quality surveys.

Filtering To improve our knowledge about w_0^de, tomography filters will be required. This allows us to extract information from different cosmic eras. Concerning mask effects, inpainting
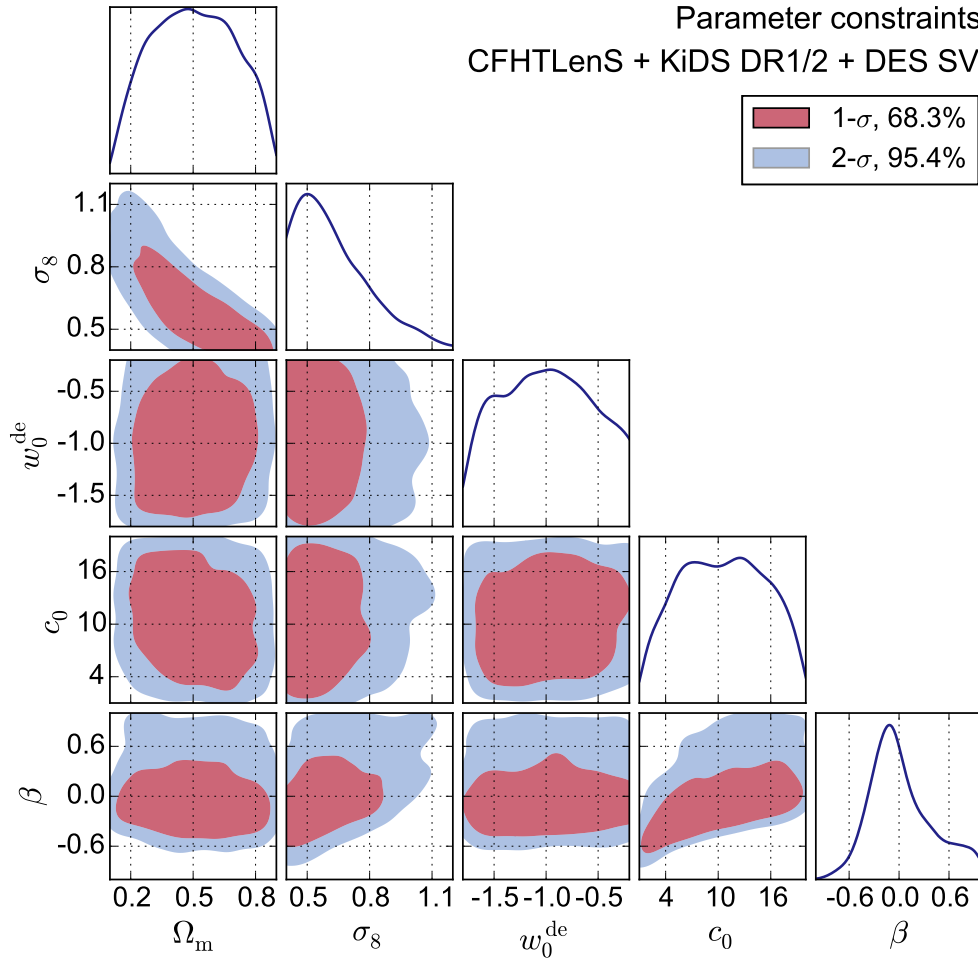
FIGURE 9.8: Preliminary result of $\Omega_{\rm m}$-σ_8-$w_0^{\rm de}$-c_0-β constraints with CFHTLenS-KiDS-DES joint data sets.

could be interesting as a tool to enhance the quantity of information. Even if the inpainted information is not included in statistics, the technique will still help to regularize mask effects and reduce the systematics.

Larger data sets Ongoing lensing surveys such as KiDS, DES, HSC, and J-PAS, and upcoming surveys such as Euclid, LSST, WFIRST all aim for providing larger data sets than CFHTLenS, considered as the current state of the art. Information from deeper and wider fields will help us refine cosmological constraints with weak-lensing peak counts.

Beyond ΛCDM Weak-lensing peak counts are sensitive to cosmological models beyond ΛCDM which change the mass function. Probing modified gravity is a future task of weak-lensing peaks. Constraints on parameters of modified gravity have recently been studied using CFHTLenS data (Liu et al. 2016). Similarly, peak counts can also be used for studying the effective number of neutrinos and the neutrino mass. The shape of the mass function changes as different models are considered.

Beyond peak counts Beyond the abundance, peaks could provide other information potentially interesting for extracting cosmological information. For example, Marian et al. (2013) estimated the Fisher information derived from tangential shear profiles of peaks and peak-peak

correlations. These are some observables to explore for future studies.

Summary

We provide constraints on Ω_m, σ_8, w_0^de, c_0, and β with data sets from CFHTLenS, KiDS DR1/2, and DES SV. This is done with our fast stochastic forward model and ABC.

We also indicate that the result can be improved by excluding bins with large number-count bias, which will be done in the near future.

Finally, possible improvements for peak-count modelling are listed. We look forward to larger statistics provided by future surveys for improving our knowledge about the Universe.

Chapter X

Conclusion

Science may be described as the art of systematic over-simplification.

— Karl Popper

This thesis studies cosmology with weak-lensing peak counts. In my work, a new model has been proposed, peak statistics has been analyzed, and cosmological constraints have been obtained.

The new model adopts a stochastic forward approach. This strategy permits to include real-world effects and guarantees flexibility. In addition, it is time-saving compared to the N-body process. These advantages make our model the ideal choice for modelling peak counts.

Starting with a mass function, our model simulates uncorrelated halos with NFW profiles. It computes the projected mass from this simulation, and adds noise to generate galaxy shape catalogues. Then it applies a mass-mapping technique to obtain a convergence map, and counts and regroups peaks by their S/N. The model is implemented by the public code Camelus.

The model is based on two hypotheses. First, diffuse and unbound matter has little impact; second, halo correlation contributes little to peak counts. This halo approach, corresponding to the one-halo term, leads to a very good approximation to peaks from N-body simulations. We examined the model in detail and discovered two compensated effects: the replacement of irregular halo profiles by the NFW ones enhances peak counts, and decorrelation decreases numbers.

In this study, the Gaussian likelihood assumption has been tested in different ways. On the one hand, the cosmological dependence of the covariance can not be neglected. Assuming a constant covariance leads to a loss of constraint power. On the other hand, the constraints from the Gaussian likelihood agree well with the results from generalized techniques: the copula likelihood and non-parametric methods such as the true likelihood and the p-value. The conclusion is that the Gaussian likelihood is a good approximation to weak-lensing peak counts.

Beyond the likelihood analysis, approximate Bayesian computation (ABC) provides an alternative for parameter constraints. It is especially competitive for problems with complex likelihood shapes. It has been shown that ABC yields a consistent result compared with the likelihood in our studied framework, and can reduce the computation time by two orders of

magnitude.

This study has also provided a framework in which filtering techniques can be fairly compared. Regarding the complexity of how cosmology is encrypted in the lensing observables, it is very difficult to determine directly an optimal filter which minimizes the contour sizes from peak counts. However, ranking filters by their performance is possible. This methodology has been followed for filter comparisons. We concluded that compensated filters are more competitive than the usual Gaussian functions and that the *separated strategy* is well more recommended than the *combined strategy* when dealing with multiscale information.

This peak-count analysis is applied to three survey data jointly: CFHTLenS, KiDS DR1/2, and DES SV. So far, preliminary results agree well with Planck, also with two-point-statistic results from CFHTLenS, KiDS, and DES. The quality of the constraints is characterized by $\Delta\Sigma_8 = 0.13$ and FoM $= 5.2$.

Various improvements for the new model are possible. How to properly model the halo correlations? How to deal with the intrinsic scatter of the halo concentration parameters? Does baryonic physics influence peak counts? How do peaks response to intrinsic alignment of the galaxies? What additional information can we extract from a tomography filter? What are the numbers of peaks in a massive neutrino cosmology and in a modified gravity model? These are the questions worth investigating for the future.

Conceptually, two ideas have been promoted and put forward by this thesis. The first one is the notion of *fast stochastic forward modelling*, which provides an approach to predict non-Gaussian observables and to account for complex survey effects; the other one is *approximate Bayesian computation*, which is a promising likelihood-free parameter inference tool. As the need for exploring higher-order statistics increases, it is expected that these two concepts rise for future cosmological studies.

Abbreviations

1D	one-dimensional
2D	two-dimensional
2PCF	two-point-correlation function
3D	three-dimensional
3PCF	three-point-correlation function
ABC	approximate Bayesian computation
AGN	active galactic nucleus
BAO	baryon acoustic oscillations
CALCLENS	Curved-sky grAvitational Lensing for Cosmological LightconE simulatioNS
CAMELUS	Counts of Amplified Mass Elevations from Lensing with Ultrafast Simulation
CCD	charge-coupled device
CDC	cosmology-dependent covariance
CDF	cumulative distribution function
CDM	cold dark matter
CFHT	Canada-France-Hawaii Telescope
CFHTLenS	Canada-France-Hawaii Telescope Lensing Survey
CFHTLS	Canada-France-Hawaii Telescope Legacy Survey
CMB	cosmic microwave background
COSEBI	complete orthogonal sets of E-/B-mode integral
COSMOS	Cosmological Evolution Survey
CPU	central processing unit
cst	constant
DES	Dark Energy Survey
DES SV	Dark Energy Survey science verification
DIM-ACAV	Domain d'Intérêt Majeur en Astrophysique et Conditions d'Apparition de la Vie
DLS	Deep Lens Survey
DM	dark matter
FASTLENS	FAst STatistics for weak Lensing
FFT	fast Fourier transform
FLRW	Friedmann-Lemaître-Robertson-Walker
FOF	friends of friends
FoM	figure of merit
FORS	FOcal Reducer Spectrograph

FSF	fast stochastic forward
FSL	Fan-Shan-Liu
GaBoDS	Garching-Bonn Deep Survey
GADGET	GAlaxies with Dark matter and Gas intEracT
GLIMPSE	Gravitational Lensing Inversion and MaPping with Sparse Estimators
GPU	graphics processing unit
HEALPix	Hierarchical Equal Area isoLatitude Pixelization
HOD	halo occupation distribution
HSC	Subaru Hyper Suprime-Cam
HST	Hubble Space Telescope
IA	intrinsic alignment
iid	independent and identically distributed
J-PAS	Javalambre Physics of the Accelerating Universe Astrophysical Survey
KDE	kernel density estimation
KiDS	Kilo-Degree Survey
KiDS DR1/2	Kilo-Degree Survey data releases 1 and 2
KS	Kaiser-Squires
LPT	Lagrangian perturbation theory
LSS	large-scale structure
LSST	Large Synoptic Survey Telescope
M-c	mass-concentration
MCMC	Markov Chain Monte Carlo
MF	Minkowski functional
Mice	Marenostrum Institut de Ciencias de l'Espai
MPI	message passing interface
MRLENS	Multi-Resolution methods for gravitational LENSing
NFW	Navarro-Frenk-White
NICAEA	NumerIcal Cosmology And lEnsing cAlculations
PDF	probability density function
photo-z	photometric redshift
PINOCCHIO	PINpointing Orbit-Crossing Collapsed HIerarchical Objects
PMC	population Monte Carlo
PNCG	Programme National de Cosmologie et Galaxies
PSF	point spread function
ROCKSTAR	Robust Overdensity Calculator using K-Space Topologically Adaptive Refinement
SCDM	standard cold dark matter
SDSS	Sloan Digital Sky Survey
SIS	singular isothermal sphere
SMC	sequential Monte Carlo
S/N	signal-to-noise ratio
SNIa	supernova of type Ia
SO	spherical overdensity

SZ	Sunyaev–Zel'dovich
VLT	Very Large Telescope
WFIRST	Wide Field Infrared Survey Telescope
WL	weak lensing
WMAP	Wilkinson Microwave Anisotropy Probe
ΛCDM	cosmological constant and cold dark matter

Notation

boldface	vector or matrix
lower case	physical coordinates and radii
upper case	comoving coordinates and radii, except for w, $f_K(w)$
$\dot{}$	time derivative
$\hat{}$	estimator
$\tilde{}$	Fourier transform
$*$	complex conjugate
$*$	convolution operator
∇	gradient operator
$\nabla_\perp$	transverse derivative operator
$\mathbb{1}$	indicator function
a	scale factor
A	amplitude of the power spectrum
A_{pix}	pixel area
A_ϵ	probability of passing the one-sample test
$\mathcal{A}$	distortion matrix
B	brightness (flux density)
c	light speed
c	additive correction for shape measurement
c	copula density
c	halo concentration
c_0	amplitude parameter for the halo M-c relation
C	copula
C	two-point-correlation function
$\boldsymbol{C}$	covariance matrix
d	dimension of a data vector
D	differential operator
D	distance function
D_+	growth factor
D_{A}	angular diameter distance
D_{H}	Hubble distance at the current time
D_ℓ	angular diameter distance between the lens and the observer
$D_{\ell\mathrm{s}}$	angular diameter distance between the lens and the source
D_{L}	luminosity distance
D_{p}	proper distance
D_{s}	angular diameter distance between the source and the observer
e_N	change of variables used in Chap. 5

E	energy
E	ratio of Hubble parameters
f	dimension of a parameter vector
f	filling factor
f	mass function, multiplicity function
f	quantity for NFW profiles
f_K	comoving transverse distance
f_{NL}	amplitude of the primordial non-Gaussianity
F	cumulative distribution function
g	lens efficiency
g	lensing reduced shear
$\boldsymbol{g}$	gravitational field
$g_{\mu\nu}$	metric
G	gravitational constant
G	dimensionless projected mass
G	Green function
$G_{\mu\nu}$	Einstein tensor
h	dimensionless Hubble parameter at the current time
H	Hubble parameter
$\boldsymbol{H}$	bandwidth matrix for kernel density estimation
H_0	Hubble parameter at the current time
I	integral
I	surface brightness (brightness density)
k	norm of the vector $\boldsymbol{k}$
$\boldsymbol{k}$	wave vector in 3D Fourier space
K	smoothed noiseless convergence
K	spacetime curvature
K_N	smoothed noisy convergence
$\boldsymbol{\ell}$	wave vector in 2D Fourier space
L	linear differential operator
L	log-likelihood function
L	luminosity
$\mathcal{L}$	likelihood function
$\ln$	natural logarithm, to base e
$\log$	common logarithm, to base 10
m	multiplicative correction for shape measurement
M	mass
$M_\star$	pivot mass
$M_\odot$	solar mass
M_{ap}	aperture mass
$M_{\max}$	upper limit of mass sampling
$M_{\min}$	lower limit of mass sampling
n	mass function, halo number density
n	noise field
n_{gal}	galaxy number density
n_{peak}	peak function, PDF of having a peak with a specific S/N
n_{s}	scalar index, power law index of the scalar field power spectrum

N	number of realizations
N	smoothed noise field
N_{gal}	number of galaxies
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{C})$	normal law with the mean $\boldsymbol{\mu}$ and the covariance $\boldsymbol{C}$
p	p-value
p	pressure
p	source redshift distribution
p_i	p-value corresponding to the i-σ significance
P	distribution in the data space
P	power spectrum
$P(\cdot|\boldsymbol{\pi})$	probability density of a stochastic model under a parameter set $\boldsymbol{\pi}$
$\mathcal{P}$	distribution in the parameter space
$\mathcal{P}(\cdot)$	prior distribution
$\mathcal{P}(\cdot|\boldsymbol{x}^{\mathrm{obs}})$	posterior distribution
P_{κ}	2D convergence power spectrum
Q	aperture mass filter in shear space
Q	number of particles
Q	second-order moment tensor
r	success ratio
r_{s}	scale radius
r_{stop}	shutoff parameter
r_{vir}	physical virial radius
$\boldsymbol{r}$	physical coordinates
R	Ricci scalar
R_{vir}	comoving quantity corresponding to the physical virial radius
$R_{\mu\nu}$	Ricci tensor
$\boldsymbol{R}$	comoving coordinates
s	summary statistic
S	source function
S	surface
t	cosmic time
T	number of iterations
$T_{\mu\nu}$	stress-energy tensor
$\mathcal{T}$	optical tidal matrix
u	percentile
U	aperture mass filter in convergence space
$\boldsymbol{u}$	velocity
$\boldsymbol{v}$	peculiar velocity
V	volume
w	comoving radial distance
w	equation of state
w_0^{de}	equation of state of dark energy
W	filter function
$W_{\boldsymbol{C}}$	Gaussian smoothing kernel with covariance $\boldsymbol{C}$
W_R	filter function with characteristic size R
x_N	change of variables used in Chap. 5
$\boldsymbol{x}$	data vector, vector of observables

$\boldsymbol{x}^{\mathrm{mod}}$	data vector from the model prediction
$\boldsymbol{x}^{\mathrm{obs}}$	data vector from the observation
z	complex number
z	redshift
z_{s}	source redshift
α	FDR parameter for MRLens
α	inner slope parameter for NFW profiles
α	slope parameter for brightness power law
α	slope parameter for Σ_8
α_{gal}	source distribution parameter
β	dragging value for Σ_8
β	mass power law index for the halo mass-concentration relation
β_{gal}	source distribution parameter
$\boldsymbol{\beta}$	unlensed angular position
γ	lensing shear
γ_*	quantity used in Chap. 5
γ_+	tangential shear
$\gamma_\times$	cross shear
δ	density contrast
δ	difference between summary statistics
δ	Dirac delta function
δ	Kronecker delta
δ_{c}	linear threshold for density contrast
Δ	overdensity
Δ^2	dimensionless power spectrum
ϵ	observed ellipticity
ϵ	tolerance level
$\epsilon^{(\mathrm{s})}$	intrinsic ellipticity
θ	change of variables used in Chap. 3
θ	distance in the angular space, norm of the vector $\boldsymbol{\theta}$
$\boldsymbol{\theta}$	2D angular position coordinates
θ_*	quantity used in Chap. 5
θ_1	angular coordinate component, aligned with right ascension
θ_2	angular coordinate component, aligned with declination
θ_{G}	Gaussian kernel size
θ_{ker}	kernel size
θ_N	change of variables used in Chap. 5
θ_{pix}	pixel width
θ_{s}	angular scale radius
θ_{vir}	angular virial radius
Θ	Heaviside step function
κ	lensing convergence
$\Delta\kappa$	difference between the true convergence and κ_{halo}
κ_0	free constant from the KS inversion
κ_1	correction for κ_{proj}
κ_{halo}	lensing convergence from the projected mass of a halo
κ_{proj}	lensing convergence from the projected mass along a line of sight

κ_n	convergence with noise
λ	affine parameter
λ	eigenvalue
λ	mass proportion within a specified range
λ	mass-sheet degeneracy
λ	threshold for filling factor
λ	wavelength
Λ	cosmological constant
μ	lensing magnification
ν	S/N of peaks
$\boldsymbol{\xi}$	physical separation vector between two geodesics
$\boldsymbol{\Xi}$	comoving separation vector between two geodesics
$\boldsymbol{\pi}$	parameter vector
$\boldsymbol{\pi}^{\mathrm{in}}$	input parameter vector
ρ	matter density
$\bar{\rho}$	background matter density
$\bar{\rho}_0$	background matter density at the current time
ρ_{crit}	critical density at the current time
ρ_{s}	central mass density
σ	density dispersion
σ_8	matter fluctuation under a spherical smoothing with radius 8 Mpc/h
σ_i	i-th order moment of a smoothed noise field
$\sigma_{\mathrm{noise}}^2$	variance of a smoothed noise field, squared zero-order moment
σ_{pix}^2	variance of noise in a pixel
σ_ϵ^2	variance of the intrinsic ellipticity dispersion, sum of both components
Σ	surface mass density
Σ_8	tilted matter fluctuation defined by Eq. (6.4) or Eq. (8.17)
$\Delta\Sigma_8$	width of the banana-shaped contour on the Ω_{m}-σ_8 plane
Σ_{crit}	critical surface mass density
τ	conformal time
ϕ	probability density of a normal distribution
ϕ	reduced Newtonian potential
φ	rotation angle
Φ	Newtonian potential
χ_f^2	probability density of the chi-squared distribution with f degrees of freedom
ψ	lensing potential
ω	weight, weight function
Ω	solid angle
Ω_{b}	dimensionless baryon density at the current time
Ω_K	dimensionless density corresponding to the space-time curvature
Ω_{m}	dimensionless matter density at the current time
Ω_{r}	dimensionless radiation density at the current time
Ω_Λ	dimensionless density corresponding to the cosmological constant for a ΛCDM model

Bibliography

Akeret J, Refregier A, Amara A et al. (2015). J. Cosmology Astropart. Phys., 8, 43 [1504.07245].

Albrecht A, Bernstein G, Cahn R et al. (2006). ArXiv Astrophysics e-prints [astro-ph/0609591].

Austen J (1813). *Pride and Prejudice.*

Bacon D J, Refregier A R, & Ellis R S (2000). MNRAS, 318, 625 [astro-ph/0003008].

Baltz E A, Marshall P, & Oguri M (2009). J. Cosmology Astropart. Phys., 1, 015 [0705.0682].

Bard D, Kratochvil J M, Chang C et al. (2013). ApJ, 774, 49 [1301.0830].

Bard D, Kratochvil J M, & Dawson W (2016). ApJ, 819, 158.

Bardeen J M, Bond J R, Kaiser N, & Szalay A S (1986). ApJ, 304, 15.

Bartelmann M (2010). Classical and Quantum Gravity, 27, 233001 [1010.3829].

Bartelmann M, King L J, & Schneider P (2001). A&A, 378, 361 [astro-ph/0103465].

Bartelmann M, Narayan R, Seitz S, & Schneider P (1996). ApJ, 464, L115 [astro-ph/9601011].

Bartelmann M, Perrotta F, & Baccigalupi C (2002). A&A, 396, 21 [astro-ph/0206507].

Baumann D, Nicolis A, Senatore L, & Zaldarriaga M (2012). J. Cosmology Astropart. Phys., 7, 051 [1004.2488].

Bayes T (1763). Philosophical Transactions of the Royal Society of London, 53, 269.

Beaumont M A, Cornuet J-M, Marin J-M, & Robert C P (2009). Biometrika, 96, 983–990 [0805.2256].

Beauvoir S (1949). *Le Deuxième Sexe (The Second Sex).*

Becker M R (2013). MNRAS, 435, 115 [1210.3069].

Beckett S (1953). *En attendant Godot (Waiting for Godot).*

Behroozi P S, Wechsler R H, & Wu H-Y (2013). ApJ, 762, 109 [1110.4372].

Benabed K, Cardoso J-F, Prunet S, & Hivon E (2009). MNRAS, 400, 219 [0901.4537].

Benítez N (2000). ApJ, 536, 571 [astro-ph/9811189].

Benjamini Y & Hochberg Y (1995). Journal of the Royal Statistical Society, Series B, 57, 289–300.

Bergé J, Amara A, & Réfrégier A (2010). ApJ, 712, 992 [0909.0529].

Berger J O, Fienberg S E, Raftery A E, & Robert C P (2010). Proceedings of the National Academy of Sciences, 107, E157.

Bernardeau F, Colombi S, Gaztañaga E, & Scoccimarro R (2002). Phys. Rep., 367, 1 [astro-ph/0112551].

Bernardeau F & Valageas P (2000). A&A, 364, 1 [astro-ph/0006270].

Bhattacharya S, Habib S, Heitmann K, & Vikhlinin A (2013). ApJ, 766, 32 [1112.5479].

Bhattacharya S, Heitmann K, White M et al. (2011). ApJ, 732, 122 [1005.2239].

Blum M G B, Nunes M A, Prangle D, & Sisson S A (2013). Statistical Science, 28, 189 [1202.3819].

Bobin J, Sureau F, Starck J-L et al. (2014). A&A, 563, A105 [1401.6016].

Bond J R & Efstathiou G (1987). MNRAS, 226, 655.

Bonnet H, Mellier Y, & Fort B (1994). ApJ, 427, L83.

Bornn L, Pillai N, Smith A, & Woodard D (2014). ArXiv e-prints [1404.6298].

Bourguignon S, Mary D, & Slezak E (2011). Selected Topics in Signal Processing, IEEE Journal of, 5,

1002.

Brainerd T G, Blandford R D, & Smail I (1996). ApJ, 466, 623 [astro-ph/9503073].

Bridle S & Abdalla F B (2007). ApJ, 655, L1 [astro-ph/0608002].

Bridle S & King L (2007). New Journal of Physics, 9, 444 [0705.0166].

Broadhurst T J, Taylor A N, & Peacock J A (1995). ApJ, 438, 49 [astro-ph/9406052].

Bryan G L & Norman M L (1998). ApJ, 495, 80 [astro-ph/9710107].

Bullock J S, Kolatt T S, Sigad Y et al. (2001). MNRAS, 321, 559 [astro-ph/9908159].

Calvino I (1979). *Se una notte d'inverno un viaggiatore (If on a winter's night a traveler)*.

Cameron E & Pettitt A N (2012). MNRAS, 425, 44 [1202.1426].

Candes E & Tao T (2006). Information Theory, IEEE Transactions on, 52, 5406.

Candès E J, Wakin M B, & Boyd S P (2008). Journal of Fourier Analysis and Applications, 14, 877.

Canizares C R (1981). Nature, 291, 620.

Cardone V F, Camera S, Mainini R et al. (2013). MNRAS, 430, 2896 [1204.3148].

Cardone V F, Camera S, Sereno M et al. (2015). A&A, 574, A141 [1409.5601].

Carrasco J J M, Hertzberg M P, & Senatore L (2012). Journal of High Energy Physics, 9, 82 [1206.2926].

Carrillo R E, McEwen J D, & Wiaux Y (2012). MNRAS, 426, 1223 [1205.3123].

Carroll L (1865). *Alice's Adventures in Wonderland*.

Casella G & Berger R L (2002). *Statistical Inference*, Duxbury.

Cervantes M (1605). *El Ingenioso Hidalgo Don Quixote de la Mancha (Don Quixote)*.

Chen Y-C, Ho S, Brinkmann J et al. (2015). ArXiv e-prints [1509.06443].

Connolly A J, Csabai I, Szalay A S et al. (1995). AJ, 110, 2655 [astro-ph/9508100].

Cooray A & Sheth R K (2002). Phys. Rep., 372, 1 [astro-ph/0206508].

Crittenden R G, Natarajan P, Pen U-L, & Theuns T (2002). ApJ, 568, 20 [astro-ph/0012336].

Csilléry K, Blum M G, Gaggiotti O E, & François O (2010). Trends in Ecology & Evolution, 25, 410 .

Dahle H (2006). ApJ, 653, 954 [astro-ph/0608480].

Dalal N, Doré O, Huterer D, & Shirokov A (2008). Phys. Rev. D, 77, 123514 [0710.4560].

Dark Energy Survey Collaboration, Abbott T, Abdalla F B et al. (2016). MNRAS, 460, 1270 [1601.00329].

Daubechies I, Defrise M, & De Mol C (2004). Communications on Pure and Applied Mathematics, 57, 1413.

De Boni C, Ettori S, Dolag K, & Moscardini L (2013). MNRAS, 428, 2921 [1205.3163].

de Jong J T A, Verdoes Kleijn G A, Boxhoorn D R et al. (2015). A&A, 582, A62 [1507.00742].

Del Moral P, Doucet A, & Jasra A (2006). Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68, 411–436.

Descartes R (1637). *Discours de la méthode (Discourse on the method)*.

Dickens C (1859). *A Tale of Two Cities*.

Diemer B & Kravtsov A V (2014). ApJ, 789, 1 [1401.1216].

Diemer B & Kravtsov A V (2015). ApJ, 799, 108 [1407.4730].

Dietrich J P, Böhnert A, Lombardi M et al. (2012). MNRAS, 419, 3547 [1103.4607].

Dietrich J P, Erben T, Lamer G et al. (2007). A&A, 470, 821 [0705.3455].

Dietrich J P & Hartlap J (2010). MNRAS, 402, 1049 [0906.3512].

Dodelson S (2003). *Modern cosmology*, Academic press.

Drovandi C C & Pettitt A N (2011). Biometrics, 67, 225.

Dubois Y, Gavazzi R, Peirani S, & Silk J (2013). MNRAS, 433, 3297 [1301.3092].

Duffy A R, Schaye J, Kay S T, & Dalla Vecchia C (2008). MNRAS, 390, L64 [0804.2486].

Duffy A R, Schaye J, Kay S T et al. (2010). MNRAS, 405, 2161 [1001.3447].

Dumas A (1844). *Les Trois Mousquetaires (The Three Musketeers).*

Dyson F W, Eddington A S, & Davidson C (1920). Philosophical Transactions of the Royal Society of London Series A, 220, 291.

Efstathiou G, Bernstein G, Tyson J A et al. (1991). ApJ, 380, L47.

Eifler T, Schneider P, & Hartlap J (2009). A&A, 502, 721 [0810.4254].

Einasto J (1965). Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87.

Einstein A (1915). Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin, 1915, 844.

Elad M & Aharon M (2006). Image Processing, IEEE Transactions on, 15, 3736.

Ellis G F R (2007). General Relativity and Gravitation, 39, 1047.

Erben T, Hildebrandt H, Miller L et al. (2013). MNRAS, 433, 2545 [1210.8156].

Etherington I (1933). The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 15, 761.

Fadili M, Starck J-L, & Murtagh F (2009). The Computer Journal, 52, 64 [http://comjnl.oxfordjournals.org/content/52/1/64.full.pdf+html].

Falco E E, Gorenstein M V, & Shapiro I I (1985). ApJ, 289, L1.

Fan Z (2007). ApJ, 669, 10 [0706.3941].

Fan Z, Shan H, & Liu J (2010). ApJ, 719, 1408 [1006.5121].

Fang W & Haiman Z (2007). Phys. Rev. D, 75, 043010 [astro-ph/0612187].

Fischer P, McKay T A, Sheldon E et al. (2000). AJ, 120, 1198 [astro-ph/9912119].

Fisher R A (1922). Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 222, 309 [http://rsta.royalsocietypublishing.org/content/222/594-604/309.full.pdf].

Frank A (1947). *Het Achterhuis (The Diary of a Young Girl).*

Fu L & Kilbinger M (2010). MNRAS, 401, 1264 [0907.0795].

Fu L, Kilbinger M, Erben T et al. (2014). MNRAS, 441, 2725 [1404.5469].

Fu L, Semboloni E, Hoekstra H et al. (2008). A&A, 479, 9 [0712.0884].

Gauss C F (1809). *Theoria motvs corporvm coelestivm in sectionibvs conicis solem ambientivm.*

Gavazzi R & Soucail G (2007). A&A, 462, 459 [astro-ph/0605591].

Górski K M, Hivon E, Banday A J et al. (2005). ApJ, 622, 759 [astro-ph/0409513].

Guio P & Achilleos N (2009). MNRAS, 398, 1254 [0906.1905].

Guth A H (1981). Phys. Rev. D, 23, 347.

Hahn C, Vakili M, Walsh K et al. (2016). ArXiv e-prints [1607.01782].

Hahn O, Porciani C, Carollo C M, & Dekel A (2007). MNRAS, 375, 489 [astro-ph/0610280].

Hamana T, Oguri M, Shirasaki M, & Sato M (2012). MNRAS, 425, 2287 [1204.6117].

Hamana T, Sakurai J, Koike M, & Miller L (2015). PASJ, 67, 34 [1503.01851].

Hamana T, Takada M, & Yoshida N (2004). MNRAS, 350, 893 [astro-ph/0310607].

Hartlap J, Simon P, & Schneider P (2007). A&A, 464, 399 [astro-ph/0608064].

Heavens A, Refregier A, & Heymans C (2000). MNRAS, 319, 649 [astro-ph/0005269].

Heisenberg L, Schäfer B M, & Bartelmann M (2011). MNRAS, 416, 3057 [1011.1559].

Hemingway E (1952). *The Old Man and the Sea.*

Hennawi J F & Spergel D N (2005). ApJ, 624, 59 [astro-ph/0404349].

Henry J P (2000). ApJ, 534, 565 [astro-ph/0002365].

Hetterscheidt M, Erben T, Schneider P et al. (2005). A&A, 442, 43 [astro-ph/0504635].

Heymans C, Van Waerbeke L, Miller L et al. (2012). MNRAS, 427, 146 [1210.0032].

Higuchi Y & Shirasaki M (2016). ArXiv e-prints [1603.01325].

Hilbert S, Marian L, Smith R E, & Desjacques V (2012). MNRAS, 426, 2870 [1204.4530].

Hildebrandt H, Viola M, Heymans C et al. (2016). ArXiv e-prints [1606.05338].

Hinshaw G, Larson D, Komatsu E et al. (2013). ApJS, 208, 19 [1212.5226].

Hirata C M & Seljak U (2004). Phys. Rev. D, 70, 063526 [astro-ph/0406275].

Iba Y (2001). Transactions of the Japanese Society for Artificial Intelligence, 16, 279 [cond-mat/0008226].

Ishida E E O, Vitenti S D P, Penna-Lima M et al. (2015). Astronomy and Computing, 13, 1 [1504.06129].

Jain B & Seljak U (1997). ApJ, 484, 560 [astro-ph/9611077].

Jain B, Seljak U, & White S (2000). ApJ, 530, 547 [astro-ph/9901191].

Jain B & Van Waerbeke L (2000). ApJ, 530, L1 [astro-ph/9910459].

Jaroszyński M & Kostrzewa Z (2010). Acta Astron., 60, 41.

Jarvis M, Sheldon E, Zuntz J et al. (2016). MNRAS[1507.05603].

Jee M J, Tyson J A, Schneider M D et al. (2013). ApJ, 765, 74 [1210.2732].

Jenkins A, Frenk C S, White S D M et al. (2001). MNRAS, 321, 372 [astro-ph/0005260].

Jennings E & Madigan M (2016). ArXiv e-prints [1608.07606].

Jennings E, Wolf R, & Sako M (2016). ArXiv e-prints [1611.03087].

Jiao Y-X, Shan H-Y, & Fan Z-H (2011). Research in Astronomy and Astrophysics, 11, 507 [1012.5894].

Jing Y P & Suto Y (2002). ApJ, 574, 538 [astro-ph/0202064].

Joudaki S, Blake C, Heymans C et al. (2016). ArXiv e-prints [1601.05786].

Jullo E, Pires S, Jauzac M, & Kneib J-P (2014). MNRAS, 437, 3969 [1309.5718].

Kacprzak T, Kirk D, Friedrich O et al. (2016). MNRAS, 463, 3653 [1603.05040].

Kafka F (1915). *Die Verwandlung (The Metamorphosis)*.

Kainulainen K & Marra V (2011). Phys. Rev. D, 83, 023009 [1011.0732].

Kaiser N (1992). ApJ, 388, 272.

Kaiser N & Squires G (1993). ApJ, 404, 441.

Kaiser N, Squires G, Fahlman G, & Woods D. (1994). in Clusters of Galaxies, Durret F, Mazure A, & Tran Thanh Van J (eds.). p. 269

Kilbinger M, Benabed K, Guy J et al. (2009). A&A, 497, 677 [0810.5129].

Kilbinger M, Fu L, Heymans C et al. (2013). MNRAS, 430, 2200 [1212.3338].

Kilbinger M & Schneider P (2005). A&A, 442, 69 [astro-ph/0505581].

Killedar M, Borgani S, Fabjan D et al. (2015). ArXiv e-prints [1507.05617].

King L J & Mead J M G (2011). MNRAS, 416, 2539 [1105.3155].

Kirk D, Brown M L, Hoekstra H et al. (2015). Space Sci. Rev., 193, 139 [1504.05465].

Kratochvil J M, Haiman Z, & May M (2010). Phys. Rev. D, 81, 043519 [0907.0486].

Kratochvil J M, Lim E A, Wang S et al. (2012). Phys. Rev. D, 85, 103513 [1109.6334].

Kristiansen J R, Elgarøy Ø, & Dahle H (2007). Phys. Rev. D, 75, 083510 [astro-ph/0611761].

Kruse G & Schneider P (1999). MNRAS, 302, 821 [astro-ph/9806071].

Kruse G & Schneider P (2000). MNRAS, 318, 321 [astro-ph/9904192].

Kuijken K, Heymans C, Hildebrandt H et al. (2015). MNRAS, 454, 3500 [1507.00738].

Kundera M (1984). *Nesnesitelná lehkost bytí (The Unbearable Lightness of Being)*.

Lambert P, Pires S, Ballot J et al. (2006). A&A, 454, 1021 [astro-ph/0604092].

Lanusse F, Starck J-L, Leonard A, & Pires S (2016). ArXiv e-prints [1603.01599].

Laureijs R, Amiaux J, Arduini S et al. (2011). ArXiv e-prints [1110.3193].

Leclercq F, Jasche J, Gil-Marín H, & Wandelt B (2013). J. Cosmology Astropart. Phys., 11, 48 [1305.4642].

Leonard A, Lanusse F, & Starck J-L (2014). MNRAS, 440, 1281 [1308.1353].

Li M (2011). Chinese Astron. Astrophys., 35, 349.

Limber D N (1953). ApJ, 117, 134.

Lin C-A & Kilbinger M (2015a). A&A, 576, A24 [1410.6955].

Lin C-A & Kilbinger M (2015b). A&A, 583, A70 [1506.01076].

Lin C-A, Kilbinger M, & Pires S (2016). A&A, 593, A88 [1603.06773].

Liu J & Haiman Z (2016). Phys. Rev. D, 94, 043533 [1606.01318].

Liu J, Haiman Z, Hui L et al. (2014). Phys. Rev. D, 89, 023515 [1310.7517].

Liu J, Petri A, Haiman Z et al. (2015a). Phys. Rev. D, 91, 063507 [1412.0757].

Liu X, Li B, Zhao G-B et al. (2016). Physical Review Letters, 117, 051101 [1607.00184].

Liu X, Pan C, Li R et al. (2015b). MNRAS, 450, 2888 [1412.3683].

Liu X, Wang Q, Pan C, & Fan Z (2014). ApJ, 784, 31 [1304.2873].

Luppino G A & Kaiser N (1997). ApJ, 475, 20 [astro-ph/9601194].

MacCrann N, Zuntz J, Bridle S et al. (2015). MNRAS, 451, 2877 [1408.4742].

Mainini R & Romano A (2014). J. Cosmology Astropart. Phys., 8, 63 [1402.0697].

Makino N, Sasaki M, & Suto Y (1992). Phys. Rev. D, 46, 585.

Maoli R, Van Waerbeke L, Mellier Y et al. (2001). A&A, 368, 766 [astro-ph/0011251].

Marian L & Bernstein G M (2006). Phys. Rev. D, 73, 123525 [astro-ph/0605746].

Marian L, Hilbert S, Smith R E et al. (2011). ApJ, 728, L13 [1010.5242].

Marian L, Smith R E, & Bernstein G M (2009). ApJ, 698, L33 [0811.1991].

Marian L, Smith R E, & Bernstein G M (2010). ApJ, 709, 286 [0912.0261].

Marian L, Smith R E, Hilbert S, & Schneider P (2012). MNRAS, 423, 1711 [1110.4635].

Marian L, Smith R E, Hilbert S, & Schneider P (2013). MNRAS, 432, 1338 [1301.5001].

Marjoram P, Molitor J, Plagnol V, & Tavaré S (2003). Proceedings of the National Academy of Sciences, 100, 15324.

Márquez G G (1967). *Cien años de soledad (One Hundred Years of Solitude)*.

Martinet N, Bartlett J G, Kiessling A, & Sartoris B (2015). A&A, 581, A101 [1506.02192].

Martizzi D, Teyssier R, Moore B, & Wentz T (2012). MNRAS, 422, 3081 [1112.2752].

Maturi M, Angrick C, Pace F, & Bartelmann M (2010). A&A, 519, A23 [0907.1849].

Maturi M, Fedeli C, & Moscardini L (2011). MNRAS, 416, 2527 [1101.4175].

Maturi M, Meneghetti M, Bartelmann M et al. (2005). A&A, 442, 851 [astro-ph/0412604].

Maturi M, Schirmer M, Meneghetti M et al. (2007). A&A, 462, 473 [astro-ph/0607254].

McKinley T, Cook A R, & Deardon R (2009). The International Journal of Biostatistics, 5, A24.

Meeds E, Leenders R, & Welling M (2015). ArXiv e-prints [1503.01916].

Meneghetti M, Rasia E, Vega J et al. (2014). ApJ, 797, 34 [1404.1384].

Merten J, Meneghetti M, Postman M et al. (2015). ApJ, 806, 4 [1404.1376].

Miller L, Heymans C, Kitching T D et al. (2013). MNRAS, 429, 2858 [1210.8201].

Miller L, Kitching T D, Heymans C et al. (2007). MNRAS, 382, 315 [0708.2340].

Miyazaki S, Hamana T, Shimasaku K et al. (2002). ApJ, 580, L97 [astro-ph/0210441].

Miyazaki S, Oguri M, Hamana T et al. (2015). ApJ, 807, 22 [1504.06974].

Monaco P, Theuns T, & Taffoni G (2002). MNRAS, 331, 587 [astro-ph/0109323].

Munshi D & Jain B (2000). MNRAS, 318, 109 [astro-ph/9911502].

Navarro J F, Frenk C S, & White S D M (1995). MNRAS, 275, 720 [astro-ph/9408069].

Navarro J F, Frenk C S, & White S D M (1996). ApJ, 462, 563 [astro-ph/9508025].

Navarro J F, Frenk C S, & White S D M (1997). ApJ, 490, 493 [astro-ph/9611107].

Newman J A (2008). ApJ, 684, 88 [0805.1409].

Ngolè Mboula F M, Starck J-L, Ronayette S et al. (2015). A&A, 575, A86 [1410.7679].

Oesch P A, Brammer G, van Dokkum P G et al. (2016). ApJ, 819, 129 [1603.00461].

Okamoto T, Shimizu I, & Yoshida N (2014). PASJ, 66, 70 [1404.7579].

Orwell G (1945). *Animal Farm: A Fairy Story.*

Osato K, Shirasaki M, & Yoshida N (2015). ApJ, 806, 186 [1501.02055].

Pace F, Maturi M, Meneghetti M et al. (2007). A&A, 471, 731 [astro-ph/0702031].

Peacock J A (1999). *Cosmological physics*, Cambridge university press.

Peebles P J E (1980). *The large-scale structure of the universe*, Princeton university press.

Peebles P J E (1993). *Principles of physical cosmology*, Princeton University Press.

Peel A, Lin C-A, Lanusse F et al. (2016). ArXiv e-prints [1612.02264].

Petri A, Haiman Z, Hui L et al. (2013). Phys. Rev. D, 88, 123002 [1309.4460].

Petri A, Liu J, Haiman Z et al. (2015). Phys. Rev. D, 91, 103511 [1503.06214].

Pires S, Leonard A, & Starck J-L (2012). MNRAS, 423, 983 [1203.2877].

Pires S, Starck J-L, Amara A et al. (2009a). A&A, 505, 969 [0904.2995].

Pires S, Starck J-L, Amara A et al. (2009b). MNRAS, 395, 1265 [0804.4068].

Planck Collaboration, Ade P A R, Aghanim N et al. (2015a). ArXiv e-prints [1502.01589].

Planck Collaboration, Ade P A R, Aghanim N et al. (2015b). ArXiv e-prints [1502.01597].

Poisson S D (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile: précédées des règles générales du calcul des probabilités*, Bachelier.

Prangle D (2016). Statistics and Computing, 26, 171 [1405.7867].

Press W H & Schechter P (1974). ApJ, 187, 425.

Pritchard J K, Seielstad M T, Perez-Lezaun A, & Feldman M W (1999). Molecular biology and evolution, 16, 1791.

Reblinsky K, Kruse G, Jain B, & Schneider P (1999). A&A, 351, 815 [astro-ph/9907250].

Reischke R, Maturi M, & Bartelmann M (2016). MNRAS, 456, 641 [1507.01953].

Riess A G, Filippenko A V, Challis P et al. (1998). AJ, 116, 1009 [astro-ph/9805201].

Roberts D H & Odell S L (1979). A&A, 76, 254.

Robin A C, Reylé C, Fliri J et al. (2014). A&A, 569, A13 [1406.5384].

Romano A, Mainini R, Maoli R, & Scaramella R (2012). Memorie della Societa Astronomica Italiana Supplementi, 19, 277.

Rubin D B (1984). Ann. Statist., 12, 1151.

Rüschendorf L (2009). Journal of Statistical Planning and Inference, 139, 3921.

Saint-Exupéry A d (1943). *Le Petit Prince (The Little Prince).*

Sato J, Takada M, Jing Y P, & Futamase T (2001). ApJ, 551, L5 [astro-ph/0104015].

Sato M, Ichiki K, & Takeuchi T T (2011). Phys. Rev. D, 83, 023501 [1011.4997].

Schaap W E. (2007). PhD thesis. Kapteyn Astronomical Institute

Scherrer R J, Berlind A A, Mao Q, & McBride C K (2010). ApJ, 708, L9 [0909.5187].

Schirmer M. (2004). PhD thesis. Rheinischen Friedrich-Wilhelms-Universitat Bonn

Schirmer M, Erben T, Hetterscheidt M, & Schneider P (2007). A&A, 462, 875 [astro-ph/0607022].

Schirmer M, Erben T, Schneider P et al. (2004). A&A, 420, 75 [astro-ph/0401203].

Schmidt F & Rozo E (2011). ApJ, 735, 119 [1009.0757].

Schmidt S J, Ménard B, Scranton R et al. (2013). MNRAS, 431, 3307 [1303.0292].

Schneider P (1996). MNRAS, 283, 837 [astro-ph/9601039].

Schneider P. (2005). In: *Gravitational Lensing: Strong, Weak, & Micro*, Meylan G, Jetzer P, & North P (eds.), Springer-Verlag, Berlin, p. 273 [astro-ph/0509252].

Schneider P, Eifler T, & Krause E (2010). A&A, 520, A116 [1002.2136].

Schneider P & Kilbinger M (2007). A&A, 462, 841 [astro-ph/0605084].

Schneider P, Kilbinger M, & Lombardi M (2005). A&A, 431, 9 [astro-ph/0308328].

Schneider P & Lombardi M (2003). A&A, 397, 809 [astro-ph/0207454].

Schneider P, Van Waerbeke L, Jain B, & Kruse G (1998). MNRAS, 296, 873 [astro-ph/9708143].

Schneider P, van Waerbeke L, & Mellier Y (2002). A&A, 389, 729 [astro-ph/0112441].

Schuhmann R L, Joachimi B, & Peiris H V (2016). MNRAS, 459, 1916 [1510.00019].

Scoccimarro R & Sheth R K (2002). MNRAS, 329, 629 [astro-ph/0106120].

Seitz C & Schneider P (1995). A&A, 297, 287 [astro-ph/9408050].

Seitz C & Schneider P (1997). A&A, 318, 687 [astro-ph/9601079].

Seitz S, Schneider P, & Ehlers J (1994). Classical and Quantum Gravity, 11, 2345 [astro-ph/9403056].

Seldner M & Peebles P J E (1979). ApJ, 227, 30.

Semboloni E, Schrabback T, Van Waerbeke L et al. (2011). MNRAS, 410, 143 [1005.4941].

Shakespeare W (1603). *Hamlet*.

Shan H, Kneib J-P, Tao C et al. (2012). ApJ, 748, 56 [1108.1981].

Shan H Y, Kneib J-P, Comparat J et al. (2014). MNRAS, 442, 2534 [1311.1319].

Shapiro C & Dodelson S (2007). Phys. Rev. D, 76, 083515 [0706.2395].

Sheth R K & Tormen G (1999). MNRAS, 308, 119 [astro-ph/9901122].

Sheth R K & Tormen G (2002). MNRAS, 329, 61 [astro-ph/0105113].

Shirasaki M (2016). ArXiv e-prints [1610.00840].

Shirasaki M, Hamana T, & Yoshida N (2016a). PASJ, 68, 4 [1508.02104].

Shirasaki M, Nishimichi T, Li B, & Higuchi Y (2016b). ArXiv e-prints [1610.03600].

Shirasaki M & Yoshida N (2014). ApJ, 786, 43 [1312.5032].

Shirasaki M, Yoshida N, Hamana T, & Nishimichi T (2012). ApJ, 760, 45 [1204.4981].

Silverman B W (1986). *Density estimation for statistics and data analysis*, Chapman & Hall.

Sisson S A, Fan Y, & Tanaka M M (2007). Proceedings of the National Academy of Sciences, 104, 1760.

Sisson S A, Fan Y, & Tanaka M M (2009). Proceedings of the National Academy of Sciences, 106, 16889 [http://www.pnas.org/content/106/39/16889.1.full.pdf].

Sklar A (1959). Publ. Inst. Statist. Univ. Paris, 8, 229.

Smith R E & Marian L (2011). MNRAS, 418, 729 [1106.1665].

Smith R E, Peacock J A, Jenkins A et al. (2003). MNRAS, 341, 1311 [astro-ph/0207664].

Spergel D N, Verde L, Peiris H V et al. (2003). ApJS, 148, 175 [astro-ph/0302209].

Springel V (2005). MNRAS, 364, 1105 [astro-ph/0505010].

Starck J-L, Murtagh F, & Fadili J M (2002). *Sparse Image and Signal Processing*, Cambridge University Press.

Starck J-L, Pires S, & Réfrégier A (2006). A&A, 451, 1139 [astro-ph/0503373].

Takada M & Bridle S (2007). New Journal of Physics, 9, 446 [0705.0163].

Takada M & Jain B (2002). MNRAS, 337, 875 [astro-ph/0205055].

Takada M & Jain B (2003a). ApJ, 583, L49 [astro-ph/0210261].

Takada M & Jain B (2003b). MNRAS, 340, 580 [astro-ph/0209167].

Takada M & Jain B (2003c). MNRAS, 344, 857 [astro-ph/0304034].

Tang J Y & Fan Z H (2005). ApJ, 635, 60 [astro-ph/0508508].

Taruya A, Takada M, Hamana T et al. (2002). ApJ, 571, 638 [astro-ph/0202090].

Tavaré S, Balding D J, Griffiths R C, & Donnelly P (1997). Genetics, 145, 505.

The Dark Energy Survey Collaboration, Abbott T, Abdalla F B et al. (2015). ArXiv e-prints [1507.05552].

Tinker J, Kravtsov A V, Klypin A et al. (2008). ApJ, 688, 709 [0803.2706].

Tolstoy L N (1877). *Anna Karenina*.

Turner E L (1980). ApJ, 242, L135.

Umetsu K, Medezinski E, Nonino M et al. (2014). ApJ, 795, 163 [1404.1375].

Van Waerbeke L (2000). MNRAS, 313, 524 [astro-ph/9909160].

Van Waerbeke L, Benjamin J, Erben T et al. (2013). MNRAS, 433, 3373 [1303.1806].

Van Waerbeke L, Mellier Y, Erben T et al. (2000). A&A, 358, 30 [astro-ph/0002500].

VanderPlas J T, Connolly A J, Jain B, & Jarvis M (2012). ApJ, 744, 180 [1109.5175].

Viola M, Cacciato M, Brouwer M et al. (2015). MNRAS, 452, 3529 [1507.00735].

Wang S, Haiman Z, & May M (2009). ApJ, 691, 547 [0809.4052].

Wang S, Khoury J, Haiman Z, & May M (2004). Phys. Rev. D, 70, 123008 [astro-ph/0406331].

Warren M S, Abazajian K, Holz D E, & Teodoro L (2006). ApJ, 646, 881 [astro-ph/0506395].

Weinberg N N & Kamionkowski M (2002). MNRAS, 337, 1269 [astro-ph/0203061].

Weinberg N N & Kamionkowski M (2003). MNRAS, 341, 251 [astro-ph/0210134].

Weyant A, Schafer C, & Wood-Vasey W M (2013). ApJ, 764, 116 [1206.2563].

White M, van Waerbeke L, & Mackey J (2002). ApJ, 575, 640 [astro-ph/0111490].

Wittman D, Dell'Antonio I P, Hughes J P et al. (2006). ApJ, 643, 128 [astro-ph/0507606].

Wittman D, Tyson J A, Margoniner V E et al. (2001). ApJ, 557, L89 [astro-ph/0104094].

Wraith D, Kilbinger M, Benabed K et al. (2009). Phys. Rev. D, 80, 023507 [0903.0837].

Wright C O & Brainerd T G (2000). ApJ, 534, 34.

Yang X, Kratochvil J M, Huffenberger K et al. (2013). Phys. Rev. D, 87, 023511 [1210.0608].

Yang X, Kratochvil J M, Wang S et al. (2011). Phys. Rev. D, 84, 043529 [1109.6333].

Yoo J & Seljak U (2012). Phys. Rev. D, 86, 083504 [1207.2471].

Zaldarriaga M & Scoccimarro R (2003). ApJ, 584, 559 [astro-ph/0208075].

Zambom A Z & Dias R (2012). ArXiv e-prints [1212.2812].

Zorrilla Matilla J M, Haiman Z, Hsu D et al. (2016). Phys. Rev. D, 94, 083506 [1609.03973].

Zwicky F (1933). Helvetica Physica Acta, 6, 110.

Index