Ray-Tracing Simulations
of
Weak Gravitational Lensing

vorgelegt von

Jan Hartlap

aus
Essen

Bonn 2009

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Peter Schneider
2. Gutachter: Prof. Dr. Andreas Eckart

Tag der Promotion: 23. Juni 2009

Contents

1 Introduction

According to Einstein's General Theory of Relativity, light is deflected by gravitational fields. This effect is omnipresent on cosmological scales: light rays emitted by distant galaxies are continuously deflected by the large scale structure of the Universe, leading to a coherent distortion of the images of the galaxies. This pattern directly encodes information about the properties of the mass distribution between the observer and the source galaxies.

Since its first detection in the 1990s, this so-called weak gravitational lens effect has become a powerful method to constrain cosmological parameters and to study the matter distribution in the Universe. It does not rely on assumptions on the relation between luminous and dark matter or the dynamical state of the objects under consideration; therefore, it constitutes an important, independent complement to other cosmological probes such as, e.g., the clustering of galaxies. In fact, this property of weak lensing makes it an ideal method to study the relation of galaxies to the underlying dark matter distribution, the so-called galaxy bias.

In the near future, several large observational campaigns such as the Kilo-Degree Survey (KIDS) and the Panoramic Survey Telescope and Rapid Response System survey (PanSTARRS) will allow measurements of weak gravitational lensing with unprecedented signal-to-noise levels. To make optimal use of these data sets, it is mandatory to have accurate theoretical predictions for the weak lensing effects for various cosmological models to which the measurements can be compared. Presently, however, this prerequisite is not yet fulfilled: modelling the non-linear gravitational evolution of the matter distribution in the Universe is difficult and can only be done using costly numerical simulations. For this reason, only fitting formulae of limited accuracy are available at present to describe the dependence of the statistical properties of the large-scale structure on the cosmological parameters. The transition from the properties of the matter distribution to analytical predictions for weak gravitational lensing requires further approximations that are not accurate enough in all cases.

In this thesis, we make a step towards increasing the quality of the theoretical predictions of the weak lensing effect. We use one of the largest and most current simulations of structure formation, the Millennium Simulation, in combination with a semi-analytic model for the formation of galaxies, to carry out ray-tracing simulations of weak gravitational lensing by the cosmological large-scale structure. The improvement of the theoretical predictions can only be translated into more accurate constraints on cosmological parameters if adequate statistical tools are available to compare theoretical results to observational data. Therefore, we complement our simulation effort with a discussion of several aspects of the statistical methods that are currently used to analyze weak lensing measurements.

In detail, this thesis is organized as follows:

- In Chapter 2, we review the cosmological framework relevant for this thesis and discuss the most important aspects of structure formation in the Universe.

- Chapter 3 presents the formalism of weak gravitational lensing. We introduce several two-point statistics of the distortion field and relate them to the properties of the matter

distribution in the Universe. Furthermore, we discuss weak gravitational lensing by galaxies and the influence of weak gravitational lensing on the angular clustering of galaxies.

- Chapter 4 contains an overview of the most important techniques for simulating the structure formation in the Universe. We also briefly describe algorithms for the identification of gravitationally bound structures, so-called halos, and semi-analytic methods to populate these halos with galaxies. Finally, we introduce the N-body simulations used for this thesis.

- In Chapter 5, we describe the multiple-lens-plane algorithm, which is the essence of ray-tracing simulations of weak gravitational lensing. We discuss two different implementations, suited for N-body simulations of different sizes. As a first application, we study the accuracy of current theoretical predictions for weak lensing two-point statistics.

- Chapter 6 is concerned with the statistical analysis of weak lensing surveys. We report on a method to obtain an unbiased estimate of the inverse covariance matrix from simulations or observational data, which is needed to compute the likelihood function for the estimation of cosmological parameters. We investigate the accuracy of the Gaussian approximation to the cosmic shear likelihood using both a large set of ray-tracing simulations and a novel method to estimate high-dimensional probability distributions. We quantify the effect of the non-Gaussianity on the constraints on cosmological parameters and apply our technique to re-analyze cosmic shear data obtained from the Chandra Deep Field South.

- In Chapter 7, we use our ray-tracing simulations to gain a deeper understanding of weak lensing by galaxies and develop a simple halo model. Furthermore, we compare our simulation results to measurements from the Sloan Digital Sky Survey.

- Chapter 8 focuses on the relation of the distribution of galaxies to that of the dark matter component. Within the framework of linear stochastic biasing, the galaxy bias is described by two scale-dependent parameters. We study how accurately these functions can be predicted from N-body simulations directly and how well they can be measured using weak lensing.

The thesis concludes with a summary and an outlook.

2 The Cosmological Standard Model

In the past decade, cosmology has seen a dramatic development. Huge data sets have become available through the advent of large-scale surveys such as the *Sloan Digital Sky Survey* (SDSS, e.g. York et al. 2000), which have greatly increased our knowledge of the distribution of galaxies, or experiments such as COBE (Smoot et al. 1991) the *Wilkinson Microwave Anisotropy Probe* (WMAP, Bennett et al. 1997), which have mapped the Cosmic Microwave Background (CMB) anisotropies with unprecedented accuracy. The remarkable agreement on the properties of our Universe achieved by all those very different observations has led to the establishment of the cosmological standard model. It states that ordinary baryonic matter, as seen mostly in the form of stars and gas, only constitutes about four percent of the total energy-matter content of the Universe. About 23% of the total energy density is in the form of non-baryonic, "dark" matter, which reveals its presence only by its gravitational interaction. The remaining $\approx 73\%$ are contributed by the mysterious dark energy, which is held responsible for the observed accelerated expansion of the Universe at late times.

In this section, we will review the basics of the cosmological standard model as far as they are relevant for this thesis. More detailed treatments can be found in, e.g., Peacock (1999), Dodelson (2003), Schneider (2006a) or Weinberg (1972, 2008), from which also most of the material in this section was drawn.

2.1 Homogeneous and isotropic world models

The theoretical foundation of the standard model is the General Theory of Relativity (Einstein 1916), which is the fundamental theory of gravity – the only known force effective on length scales relevant for cosmology. Together with the Cosmological Principle, it determines the geometry and dynamics of the universe of the standard model.

2.1.1 The field equations of General Relativity

In General Relativity, gravity is an effect of the geometrical properties of the *space-time*, a $(3 + 1)$-dimensional Riemannian manifold. Its geometry is described by the symmetric metric tensor $g_{\mu\nu}$, which obeys the field equations of General Relativity:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = -\frac{8\pi G}{c^4} T_{\mu\nu} \ . \tag{2.1}$$

Here, the Einstein tensor $G_{\mu\nu}$ is a function of the metric tensor only and encodes the geometry of space-time. On the other side of the equation is the energy-momentum tensor, which specifies the energy and matter content of the Universe. Finally, Λ denotes the cosmological constant. Originally introduced into the field equations by Einstein to allow for a static solution, it has been revived in recent times by the discovery of the accelerated expansion of the Universe.

The presence of matter changes the geometry of space-time. Since particles travel on geodesics of the space-time manifold, the presence of a massive body will alter the trajectories compared to empty space. In the Newtonian limit this is interpreted as the effect of a gravitational force.

2.1.2 The Friedmann-Lemaitre-Robertson-Walker model

The Cosmological Principle states that the Universe is homogeneous and isotropic for a certain class of observers. These are the so-called *fundamental observers*, which are free-falling and therefore comoving with the mean motion of matter in the Universe. The claim of isotropy has been verified by observations of the large-scale distribution of galaxies and radio sources, as well as by CMB experiments (Smoot et al. 1991). The step from the observed isotropy to homogeneity is then made by assuming that the Earth is not located at a special place. This means that any fundamental observer sees an isotropic Universe, which in turn implies homogeneity.

The high symmetry of the space-time imposed by the Cosmological Principle simplifies the solution of the field equations (2.1) considerably. It has been shown by Robertson (1935) and Walker (1936) that the form of the line-element $\mathrm{d}s = g_{\mu\nu} \, \mathrm{d}x^\mu \, \mathrm{d}x^\nu$, which is the space-time interval between two events separated by the infinitesimally small vector $\mathrm{d}x^\mu$, is constrained to be of the form

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t) \left[\mathrm{d}w^2 + f_K^2(w) \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2 \right) \right] \quad . \tag{2.2}$$

The spatial coordinates are given by the *comoving radial distance* w and the angular coordinates θ and ϕ. The function $a(t)$ is called the *scale factor*, and it describes the global expansion of the Universe. By definition, $a(t_0) = 1$ at the present time. The function f_K depends on the curvature of the Universe, parametrized by the curvature constant K:

$$f_K(w) = \begin{cases} 1/\sqrt{K} \sin(\sqrt{K}w) & K > 0 \\ w & K = 0 \\ 1/\sqrt{-K} \sinh(\sqrt{-K}w) & K < 0 \end{cases} \tag{2.3}$$

It can be shown (Weinberg 1972) that isotropy and homogeneity require the energy-momentum tensor to take the perfect fluid form:

$$T_{\mu\nu} = \left(\rho(t) + \frac{p(t)}{c^2} \right) u_\mu u_\nu - p(t) \, g_{\mu\nu} \quad . \tag{2.4}$$

Here, u_μ is the four-velocity, $\rho(t)$ is the density and $p(t)$ is the pressure of the fluid. The fluid is characterized by its *equation of state*

$$p(t) = w_{\mathrm{eos}} \, \rho(t)c^2 \quad . \tag{2.5}$$

With the Robertson-Walker metric (Eq. 2.2) and Eq. (2.4) for the energy-momentum tensor, the field equations (2.1) reduce to two independent equations which, together with the equation of state (2.5), completely determine $a(t)$, $p(t)$ and $\rho(t)$:

$$\left(\frac{\dot{a}(t)}{a(t)} \right)^2 = \frac{8\pi G}{3}\rho(t) - \frac{Kc^2}{a^2(t)} + \frac{\Lambda}{3} \quad , \tag{2.6}$$

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3} \left(\rho(t) + \frac{3p(t)}{c^2} \right) + \frac{\Lambda}{3} \quad . \tag{2.7}$$

From these *Friedmann equations*, the *adiabatic equation* can be obtained:

$$\frac{\mathrm{d}(a^3\rho c^2)}{\mathrm{d}t} + p\frac{\mathrm{d}a^3}{\mathrm{d}t} = 0 \quad . \tag{2.8}$$

This relation allows one to determine dependence of the density of the various matter components on the scale factor. With the ansatz $\rho \propto a^\nu$ and Eq. (2.5), the result is

$$\rho(a) = \rho_0 a^{-3(w_{\text{eos}}+1)} \quad . \tag{2.9}$$

For nonrelativistic particles, for which the pressure is essentially zero ("dust"), and which are usually referred to as the matter component, one obtains $\rho_m \propto a^{-3}$. All relativistic particle species like photons or neutrinos are subsumed under the term radiation, and their density evolves as $\rho_r \propto a^{-4}$, since $p = \rho c^2/3$. Finally, from Eq. (2.6) it can be seen that $\Lambda \neq 0$ corresponds to a constant density of $\rho_\Lambda = \frac{\Lambda}{8\pi G}$. From Eq. (2.5) it follows directly that such a contribution would lead to $p_\Lambda = -\rho_\Lambda c^2$.

Eq. (2.6) allows the determination of the curvature constant. First, let $K = 0$, then it follows that

$$\rho(t) = \frac{3H(t)^2}{8\pi G} \equiv \rho_{\text{cr}}(t) \quad , \tag{2.10}$$

where the *critical density* $\rho_{\text{cr}}(t)$ and the *Hubble parameter* $H(t) = \dot{a}/a$ were defined. The Λ-term was absorbed into the density, so that $\rho(t) = \rho_m(t) + \rho_r(t) + \rho_\Lambda + ...$ The critical density defines a natural density scale, corresponding to a universe with a flat geometry. It therefore makes sense to define the *density parameter* for species i:

$$\Omega_i(t) \equiv \frac{\rho_i(t)}{\rho_{\text{cr}}(t)} = \frac{8\pi G \rho_i(t)}{3H^2(t)} \quad . \tag{2.11}$$

Furthermore, the total density parameter is given by $\Omega_0 \equiv \sum_i \Omega_i$. Usually, these parameters are given for the present time, in which case the argument is simply omitted. Going back to the Friedmann-Eq. (2.6) and evaluating it at $t = t_0$, one finds with the definition of the *Hubble constant* $H_0 = H(t_0)$ that

$$Kc^2 = H_0^2(\Omega_0 - 1) \quad . \tag{2.12}$$

Thus, the total density determines the geometry of the Universe (see Eqns. 2.2 and 2.3). Universes with $\Omega_0 > 1 \Leftrightarrow K > 0$ are called closed, such with $\Omega_0 = 1 \Leftrightarrow K = 0$ flat, and universes with $\Omega_0 < 1 \Leftrightarrow K < 0$ open.

With these definitions, Friedmann's equation can be rewritten as

$$H^2(t) = H_0^2 \left(\frac{\Omega_{\text{r}}}{a^4} + \frac{\Omega_{\text{m}}}{a^3} + \frac{1 - \Omega_0}{a^2} + \Omega_\Lambda \right) \quad . \tag{2.13}$$

From this it can be seen that in universes in which a increases with time, cosmic evolution at early times ($a \ll 1$) must have been totally dominated by radiation. Later, the radiation contribution diminishes and other components successively come to dominate the expansion. In the standard model (see Tab. 2.1 for a recent compilation of the relevant model parameters), the radiation era is followed by a matter-dominated epoch. The transition between these two is determined by the time at which matter and radiation have equal density parameters. The corresponding scale factor is

$$a_{\text{eq}} = \frac{\Omega_{\text{r}}}{\Omega_{\text{m}}} \approx 3.2 \times 10^{-5} \, \Omega_{\text{m}}^{-1} h^{-2} \quad , \tag{2.14}$$
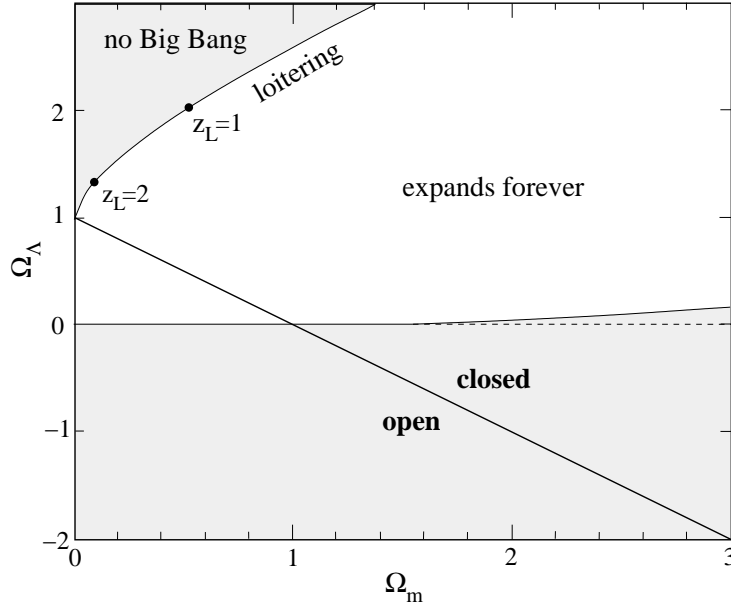
Figure 2.1: *Expansion behaviour for different density parameters (following Peacock 1999). Combinations of matter and dark energy density that yield a flat universe are given by the thick diagonal line. The maximum redshift in loitering models is denoted by z_L.*

where h is defined by $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

From Eq. (2.13), the future of a universe that has $H_0 > 0$ – like ours – can be determined. For the case that the radiation density has become negligible, the dependence of the fate of the Universe on matter and vacuum energy density can be read off from Fig. 2.1. All of these models require an initial singularity ($a(0) = 0$), except in the case of a large Ω_Λ and small matter density. In the special case of a *loitering* universe, the scale factor attains a finite minimum value in the past (corresponding to the maximum redshift z_L in Fig. 2.1). Given reasonable values for Ω_m, loitering models can be ruled out strongly, because objects at redshifts as high as $z \approx 7$ have been observed – far beyond the maximum possible redshift predicted by these models.

2.1.3 Properties of FLRW models

Hubble Law

Comoving coordinates $\boldsymbol{x}$ are related to physical or proper (spatial) coordinates $\boldsymbol{r}$ by

$$\boldsymbol{x} = \boldsymbol{r}/a(t) \ . \tag{2.15}$$

Considering a galaxy located at the origin, and another galaxy separated from the first one by $\boldsymbol{r}$, the time evolution of the separation vector is given by

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \dot{a}(t)\boldsymbol{x} = \frac{\dot{a}(t)}{a(t)}\boldsymbol{r} \ . \tag{2.16}$$

With the Hubble parameter $H(t) = \frac{\dot{a}(t)}{a(t)}$ and the Hubble constant $H_0 = H(t_0)$, one directly obtains the *Hubble Law* for the expansion in the local Universe ($t \approx t_0$):

$$\boldsymbol{u}(r) = H_0\boldsymbol{r} \ . \tag{2.17}$$

Redshift

Light rays travel along null geodesics, i.e. $ds^2 = 0$. For a radial light ray one has in addition $d\Omega = 0$. Eq. (2.2) gives the relation between comoving radial distance and the cosmic time t when the photon arriving at the observer at t_0 was emitted:

$$c\,dt = -a(t)\,dw \qquad \Rightarrow \qquad w = \int_t^{t_0} \frac{c\,dt}{a(t)} \quad . \tag{2.18}$$

The negative sign in Eq. (2.18) was chosen because on the photon path to the observer at $w = 0$, $dt > 0$, but $dw < 0$. Now let a source at distance w emit two photons at t and $t + dt_{\rm em}$, which the observer detects at t_0 and $t_0 + dt_{\rm obs}$. Employing Eq. (2.18) for both photons, it follows that $dt_{\rm em} = a(t)\,dt_{\rm obs}$, or expressed in frequencies:

$$\frac{\nu_{\rm em}}{\nu_{\rm obs}} = \frac{1}{a(t)} \equiv 1 + z \quad , \tag{2.19}$$

where the *redshift* z was defined.

Distance measures

In a curved and expanding universe, the notion of "distance" is not a well-defined concept. In fact, the distance to an object depends on the measurement method used. Two important ways of distance determination in cosmology are to measure the apparent size of a *standard ruler* (an object of known physical size) or the apparent brightness of a *standard candle* (an object of known luminosity). While in an Euclidean space both methods yield the same result, they differ in a curved Robertson-Walker space-time.

Consider two light rays emitted simultaneously from the edges of a source with diameter dl, which arrive at the observer enclosing an angle $d\theta$. The *angular diameter distance* is defined as

$$D_< \equiv \frac{dl}{d\theta} = a(t)\,f_K(w) \quad , \tag{2.20}$$

where the last equality can be read off from the metric (Eq. 2.2).

The method using a standard candle requires knowledge of the luminosity L of a given source. Assuming that the source radiates isotropically, its photons have spread out over an area $4\pi f_K^2(w)$ when arriving at the observer, are redshifted by a factor $(1+z)^{-1}$ and their arrival frequency is reduced by the same factor. The flux measured by the observer is then

$$F = \frac{L}{4\pi(1+z)^2 f_K^2(w)} \quad . \tag{2.21}$$

The *luminosity distance* is defined analogously to Euclidean space as

$$D_L \equiv \sqrt{\frac{L}{4\pi F}} = (1+z)f_K(w) = (1+z)^2 D_< \quad . \tag{2.22}$$

The horizon

Since the Big Bang, photons can have travelled only a finite distance, which is called the horizon $d_{\rm h}$. Its physical size at given cosmic time t or scale factor a can be computed from $c\,dt = -a\,dw$:

$$d_{\rm h} = a(t) \int_0^t \frac{c\,dt}{a(t)} = a(t) \int_0^a \frac{c\,da}{a^2 H(a)} \quad . \tag{2.23}$$

Regions which are separated by a distance larger than the horizon size can not have been in causal contact. However, this may change at later times as the horizon size increases.

2.2 Structure formation

The FLRW model accounts for the overall evolution of a homogeneous, isotropic universe. This is a good description for our Universe when smoothed out on scales larger than $\approx 200\,\mathrm{Mpc}$. However, there is a wealth of structures on small to intermediate length scales, such as galaxies, galaxy clusters and filaments. While it is generally being assumed that the formation of structures does not notably influence the overall expansion history of the smooth Universe, the statistical properties of the large-scale structure depend strongly on the background cosmology. In this section, we review some basic aspects of structure formation, focusing mainly on the evolution of the dark matter component.

2.2.1 The origin of structure

The origin of the small deviations from a homogeneous density which developed into the large-scale structure observed today is currently believed to be a phase of exponential expansion ("*Inflation*", see e.g. Linde 2005, and references therein) shortly after the Big Bang. In the simplest scenario, this expansion is driven by a hypothetical scalar field. During this period, initially subhorizon-sized quantum fluctuations of this field were blown up to superhorizon sizes by the rapid expansion of the Universe and therefore were imprinted permanently to the density field. At the end of the inflationary period, these small perturbations start to grow and so form the seeds of the structures observed today.

2.2.2 Linear perturbation theory

Structure growth can be roughly divided into three regimes. Shortly after the end of inflation, essentially all density fluctuations are larger than the horizon size at that time, and therefore require a treatment within the framework of General Relativity. At this time, however, the perturbations are very small, so that linear perturbation theory is sufficient to describe their evolution. As time goes by, fluctuations on larger scales enter the horizon, then allowing for a simpler linear Newtonian description. Eventually small-scale perturbations will grow so large that linear theory fails. This regime is difficult to describe analytically, and therefore is usually studied using N-body simulations.

Newtonian theory

First, we will focus on perturbations that are well inside the horizon during the matter dominated era. In this case, Newtonian physics can be used to study the evolution of these structures. Observations indicate that dark matter is non-relativistic and collisionless, so that the evolution of its phase-space density $f(\boldsymbol{r}, \boldsymbol{u})$ is determined by the collisionless Boltzmann (or Vlasov) equation

$$\frac{\partial f}{\partial t} + \boldsymbol{u}\frac{\partial f}{\partial \boldsymbol{r}} - \boldsymbol{\nabla}\Phi\frac{\partial f}{\partial \boldsymbol{u}} = 0 \tag{2.24}$$

and the Poisson equation

$$\nabla^2\phi = 4\pi G\,\rho(\boldsymbol{r}, t) - \Lambda \ \ . \tag{2.25}$$

Here, $\boldsymbol{r}$ and $\boldsymbol{u}$ are the physical position and velocity, respectively, and ϕ is the gravitational potential. The dark matter density is related to the phase space density by $\rho(\boldsymbol{x}) = m_{\mathrm{DM}} \int \mathrm{d}\boldsymbol{u}\, f(\boldsymbol{x}, \boldsymbol{u})$, where m_{DM} is the mass of a dark matter particle. This system of equations is highly non-linear, and no analytic solution is known. However, on large scales or at early times, it is possible to treat the dark matter as an ideal fluid with a unique velocity at each point in space. This is a good approximation if the density fluctuations are small, so that the dark matter particles mainly follow the mean flow determined by the large-scale gravitational potential. It breaks down when high density regions form, where particle trajectories cross frequently and no unique velocity field exists.

In the fluid approximation, the evolution of density, velocity field and gravitational potential are governed by the three coupled equations

$$\frac{\partial \rho(\boldsymbol{r}, t)}{\partial t} + \boldsymbol{\nabla}_r \cdot [\rho(\boldsymbol{r}, t) \bar{\boldsymbol{u}}(\boldsymbol{r}, t)] = 0 \qquad \text{(Continuity Eq.)}, \tag{2.26}$$

$$\frac{\partial \bar{\boldsymbol{u}}(\boldsymbol{r}, t)}{\partial t} + [\bar{\boldsymbol{u}}(\boldsymbol{r}, t) \cdot \boldsymbol{\nabla}_r]\, \bar{\boldsymbol{u}}(\boldsymbol{r}, t) = -\boldsymbol{\nabla}_r \phi(\boldsymbol{r}, t) \qquad \text{(Euler Eq.)}, \tag{2.27}$$

$$\boldsymbol{\nabla}_r^2 \phi(\boldsymbol{r}, t) = 4\pi G \rho(\boldsymbol{r}, t) - \Lambda \qquad \text{(Poisson Eq.)} \quad . \tag{2.28}$$

Here, $\bar{\boldsymbol{u}} = \int \mathrm{d}^3 u\, f(\boldsymbol{x}, \boldsymbol{u})\, \boldsymbol{u} / \int \mathrm{d}^3 u\, f(\boldsymbol{x}, \boldsymbol{u})$ is the mean velocity within the volume element $\mathrm{d}^3 x$. Eqs. (2.26) and (2.27) can be derived from the Boltzmann equation by taking moments of the phase space density. The continuity equation follows from Eq. (2.24) by intergrating over the velocity, whereas the Euler equation results from multiplication of Eq. (2.24) with $\boldsymbol{u}$ followed by velocity integration, making the additional assumption that dark matter is *cold*, i.e. that its velocity dispersion vanishes.

These equations can be expressed in terms of comoving coordinates

$$\boldsymbol{x} = \frac{\boldsymbol{r}}{a(t)} \quad , \tag{2.29}$$

$$\rho(\boldsymbol{r}, t) = \hat{\rho}\left(\frac{\boldsymbol{r}}{a(t)}, t\right) \quad , \tag{2.30}$$

$$\bar{\boldsymbol{u}}(\boldsymbol{r}, t) = \frac{\dot{a}(t)}{a(t)} \boldsymbol{r} + \boldsymbol{v}\left(\frac{\boldsymbol{r}}{a(t)}, t\right) \quad , \tag{2.31}$$

where in the last equation the velocity field was decomposed into the *Hubble flow* and the *peculiar velocity*. It is useful to introduce quantities that describe the deviation from a homogeneous Universe: the *density contrast* is defined by

$$\delta(\boldsymbol{x}, t) = \frac{\hat{\rho}(\boldsymbol{x}, t) - \bar{\rho}}{\bar{\rho}} \tag{2.32}$$

and the *comoving gravitational potential* by

$$\Phi(\boldsymbol{x}, t) = \phi[a(t)\boldsymbol{x}, t] + \frac{\ddot{a}a}{2}|\boldsymbol{x}|^2 \quad . \tag{2.33}$$

Substituting all these into Eqns. (2.26)-(2.28), one eventually obtains

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \boldsymbol{\nabla}_x \cdot [(1 + \delta)\boldsymbol{v}] = 0, \tag{2.34}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + \frac{\dot{a}}{a}\boldsymbol{v} + \frac{1}{a}\left(\boldsymbol{v} \cdot \boldsymbol{\nabla}_x\right)\boldsymbol{v} = -\frac{1}{a}\boldsymbol{\nabla}_x \Phi, \tag{2.35}$$

$$\boldsymbol{\nabla}_x^2 \Phi = \frac{3H_0^2 \Omega_{\mathrm{m}}}{2a} \delta \quad . \tag{2.36}$$

For $\delta \ll 1$, these equations can be simplified considerably by neglecting all terms of higher than linear order in δ and $\boldsymbol{v}$. This leaves the Poisson equation (2.36) unchanged, while the continuity and Euler equations become

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \boldsymbol{\nabla}_x \cdot \boldsymbol{v} = 0 \tag{2.37}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + \frac{\dot{a}}{a} \boldsymbol{v} = -\frac{1}{a} \boldsymbol{\nabla}_x \Phi \ . \tag{2.38}$$

Finally, taking the divergence of Eq. (2.38) and the time derivative of Eq. (2.37) and replacing Φ using Eq. (2.36), one obtains

$$\frac{\partial^2 \delta}{\partial t^2} + \frac{2\dot{a}}{a} \frac{\partial \delta}{\partial t} - \frac{3H_0^2 \Omega_{\mathrm{m}}}{2a^3} \delta = 0 \ . \tag{2.39}$$

Since there is no explicit spatial dependence in Eq. (2.39), every solution can be written as

$$\delta(\boldsymbol{x}, t) = D_+(t) \Delta_+(\boldsymbol{x}) + D_-(t) \Delta_-(\boldsymbol{x}) \ . \tag{2.40}$$

This means that in the linear regime the shape of a given perturbation remains unchanged in comoving coordinates, only the density contrast de- or increases. One solution of Eq. (2.39) is the Hubble parameter $H(t) = D_-(t)$, which for an expanding, matter-dominated Universe is a monotonically decreasing function of time. Any contribution of this solution to δ in Eq. (2.40) will die out quickly, and is therefore irrelevant for structure formation. The second solution can be expressed through the first one using the Wronski determinant and is given by

$$D_+(t) = \mathrm{const.} \times H_0^2 \, H(t) \int_0^t \frac{\mathrm{d}t'}{a^2(t')H^2(t')} \ . \tag{2.41}$$

The function D_+, constituting the growing mode, is called the the *growth factor* and is normalized such that it is unity at the present time. For the special case of an Einstein-de-Sitter universe ($\Omega_0 = \Omega_{\mathrm{m}} = 1$), the exact solution $D_+(t) = a(t)$ can be found.

The evolution of a dark matter perturbation depends on the cosmological epoch. If one considers a dark matter perturbation in a radiation-dominated background (neglecting curvature and dark energy contributions to $H(t)$), one can show from Eq. (2.39) that

$$D_+(a) \propto \frac{a}{a_{\mathrm{eq}}} + \frac{2}{3} \ . \tag{2.42}$$

The density contrast is basically constant while $a \ll a_{\mathrm{eq}}$; structure growth in this era is suppressed by the expansion of the Universe. Only when the matter density starts to dominate the cosmic expansion, perturbations in the dark matter density start to grow with $D_+ \propto a$.

Suppression of structure growth

If a perturbation is of a size comparable to or larger than the horizon size, the Newtonian description breaks down. One finds from a fully relativistic treatment that $\delta \propto a^2$ while radiation is the dominant species, and $\delta \propto a$ during the matter dominated era. Together with the results of linear Newtonian theory, this means that there is a characteristic scale in structure growth, namely the horizon size at matter-radiation equality, corresponding to a length scale of $d_{\mathrm{H}}(a_{\mathrm{eq}}) \approx 16(\Omega_{\mathrm{m}}h)^{-2}$ Mpc: the growth of perturbations smaller than this length scale is suppressed because
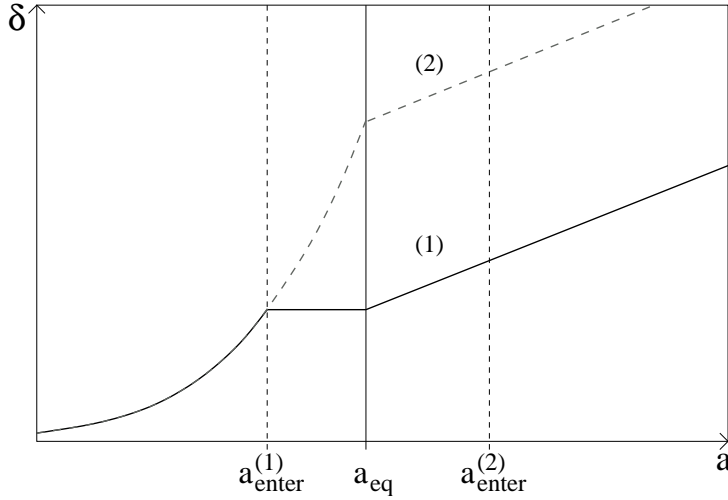
Figure 2.2: *Evolution of two perturbations; the small-scale perturbation (1) enters the horizon at $a_{\mathrm{enter}}^{(1)} < a_{\mathrm{eq}}$, the large-scale perturbation (2) enters the horizon at $a_{\mathrm{enter}}^{(2)} > a_{\mathrm{eq}}$.*

structure growth stalls for sub-horizon fluctuations during the radiation dominated era. From Fig. 2.2, it can be seen that a perturbation which enters the horizon at $a_{\mathrm{enter}}^{(1)} < a_{\mathrm{eq}}$ does not grow until the time of matter-radiation-equality. In contrast, a larger-scale perturbation that is still outside the horizon long after a_{eq} can grow continuously. Therefore, at late times it is denser by a factor of $\left(a_{\mathrm{eq}}/a_{\mathrm{enter}}^{(1)} \right)^2$.

Fourier analysis of density perturbations

It is often useful to think of the density contrast as being composed of perturbations to which a particular length scale can be assigned. In the case of a flat geometry and if one works in comoving coordinates, a decomposition of δ into Fourier modes is a natural way to achieve this. In other geometries, the Fourier modes do not form a complete set of basis functions. However, this difference only becomes important on scales comparable to the horizon size and therefore is normally neglected (Peacock 1999). Throughout, we will denote the Fourier transform of a quantity f with $\tilde{f}$. We illustrate our convention for the Fourier transform with the density contrast:

$$\delta(\boldsymbol{x}, t) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\, \mathrm{e}^{\mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}\, \tilde{\delta}(\boldsymbol{k}, t) \,, \tag{2.43}$$

where $\boldsymbol{k}$ is the comoving wave-vector.

Rewriting the linear evolution equations (Eqs. 2.36–2.38) in Fourier space, we find

$$\frac{\partial \tilde{\delta}}{\partial t} + \frac{\mathrm{i}}{a} \tilde{\boldsymbol{v}} \cdot \boldsymbol{k} = 0 \,, \tag{2.44}$$

$$\frac{\partial \tilde{\boldsymbol{v}}}{\partial t} + \frac{\dot{a}}{a} \tilde{\boldsymbol{v}} = \frac{-\mathrm{i} \boldsymbol{k}}{a} \tilde{\Phi} \,, \tag{2.45}$$

$$-k^2 \Phi = \frac{3 H_0^2 \Omega_{\mathrm{m}}}{2a} \tilde{\delta} \,. \tag{2.46}$$

This shows another benefit of working in Fourier space: in the linear regime, all Fourier modes of density and velocity evolve independently of each other. This ceases to be true when δ approaches unity and one has to expand Eqs. (2.34)–(2.36) to higher orders in δ.

Transfer functions

The picture of structure formation we have given so far is strongly simplified: besides the a suppression of small-scale perturbations due to their horizon-entry while the Universe was dominated by radiation, several other effects exist that lead to a scale-dependence of the growth of structure. For example, small-scale perturbations can be erased as soon as they enter the horizon due to fast particles streaming out of the shallower potential wells. Depending on the nature of the dark matter particles, this effect is more or less pronounced. If dark matter consists of light particles that are relativistic for quite a long time (*hot dark matter*), only the largest perturbations will survive this so-called *free streaming*. To account for these complications, one compares perturbations with wavenumber $\boldsymbol{k}$ to a reference perturbation with wavenumber $\boldsymbol{k}_{\mathrm{r}}$ that enters the horizon well after a_{eq}. At a very early time t_{i}, all fluctuations of interest are larger than the horizon. The *transfer function* $T(k)$ connects the ratio of the fluctuation amplitudes at early times to the amplitude ratio today, taking into account the scale-dependent effects of causal physics:

$$\frac{\tilde{\delta}(\boldsymbol{k}, t_0)}{\tilde{\delta}(\boldsymbol{k}_{\mathrm{r}}, t_0)} = T(k) \frac{\tilde{\delta}(\boldsymbol{k}, t_i)}{\tilde{\delta}(\boldsymbol{k}_{\mathrm{r}}, t_i)} \ . \tag{2.47}$$

An approximation to the transfer function can be obtained by the following argument: large-scale fluctuations enter the horizon long after matter-radiation equality and therefore grow unhindered. Therefore, $T(k)$ should approach unity for sufficiently small k. One finds from Eq. (2.23) for the comoving horizon size during the radiation-dominated era $x_{\mathrm{h}}(a) \propto a$. A fluctuation with comoving wavelength λ enters the horizon when $\lambda \sim 1/k \sim x_{\mathrm{h}}(a_{\mathrm{enter}})$; therefore, $a_{\mathrm{enter}} \propto 1/k$. Since the density contrast of such a fluctuation is suppressed by a factor of $(a_{\mathrm{enter}}/a_{\mathrm{eq}})^2$, the transfer function should decline asymptotically $\propto k^{-2}$. Thus,

$$T_k = \begin{cases} 1 & \text{for } 1/k \gg x_{\mathrm{h}}(a_{\mathrm{eq}}) \\ [k \, x_{\mathrm{h}}(a_{\mathrm{eq}})]^{-2} & \text{for } 1/k \ll x_{\mathrm{h}}(a_{\mathrm{eq}}) \ . \end{cases} \tag{2.48}$$

More accurate fitting functions have been given in Bardeen et al. (1986) or Eisenstein & Hu (1999), who also include corrections due to baryonic effects.

2.2.3 Statistics of the density field

No model of structure formation will be able to predict the exact density field in the Universe as we observe it today, mainly because the initial conditions are subject to stochasticity and therefore unknown. However, inflationary scenarios predict the statistical properties of the initial density field. The main goal of the theory of structure formation therefore can only be to give a description of these properties and their evolution.

Stochastic processes and random fields

The density field in the Universe can be thought of as a particular realization of an underlying *stochastic process* with certain statistical properties that are determined by the cosmological parameters. A stochastic process with index space T is a collection of random variables $R = \{Y(\boldsymbol{t}), \boldsymbol{t} \in T\}$. A particular *realization* of R is denoted by $\{y(\boldsymbol{t}), \boldsymbol{t} \in T\}$. The properties of a stochastic process can be described by its *finite-dimensional distribution functions*

$$p_{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_m}(y_1, \ldots, y_m) = \mathrm{Prob}\left[Y(\boldsymbol{t}_1) = y_1, \ldots, Y(\boldsymbol{t}_m) = y_m\right] \qquad \text{for } m \geq 1 \ , \tag{2.49}$$

which give the probability of occurrence of a specific realization of the process. Another possibility to characterize R is by its moments. The m-th moment of a stochastic process is defined by

$$\langle Y(\boldsymbol{t}_1) \cdots Y(\boldsymbol{t}_m) \rangle = \int \mathrm{d}y_1 \cdots \mathrm{d}y_m \; p_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_m}(y_1,\ldots,y_m) \, y_1 \cdots y_m \; . \tag{2.50}$$

The first moment is the mean, higher moments are called the *m-point correlation functions*. The brackets $\langle \cdot \rangle$ denote the *ensemble average*, which is the average over many realizations of R.

If $T = \mathbb{R}^n$, the stochastic process is called a *random field*, in which case we shall denote the members of the index space with $\boldsymbol{x}$. A random field is said to be *homogeneous*, if all of its finite-dimensional distribution functions are invariant under simultaneous translations. Of particular interest is the *two-point correlation function* $\xi_Y \equiv \langle Y(\boldsymbol{x}_1)Y(\boldsymbol{x}_2) \rangle$, which then only depends on $\boldsymbol{x}_1 - \boldsymbol{x}_2$. The field is called *homogeneous and isotropic*, if its finite-dimensional distribution functions are additionally invariant under simultaneous rotations. In this case, the two-point correlation function $\xi_Y \equiv \langle Y(\boldsymbol{x}_1)Y(\boldsymbol{x}_2) \rangle$ depends only on $|\boldsymbol{x}_1 - \boldsymbol{x}_2|$. A *Gaussian random field* is a random field whose finite-dimensional distribution functions are multivariate Gaussian distributions. Therefore, a Gaussian random field is fully specified by its first and second moments.

It proves to be convenient to decompose a homogeneous and isotropic random field into its Fourier components,

$$Y(\boldsymbol{x}) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \, \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \, \tilde{Y}(\boldsymbol{k}) \; . \tag{2.51}$$

Computing the Fourier-space correlation function, one finds that

$$\begin{aligned}
\left\langle \tilde{Y}(\boldsymbol{k}) \, \tilde{Y}^*(\boldsymbol{k}') \right\rangle &= (2\pi)^3 \, \delta_\mathrm{D}(\boldsymbol{k} - \boldsymbol{k}') \int \mathrm{d}^3 x' \, \mathrm{e}^{-\mathrm{i}\boldsymbol{x}'\cdot\boldsymbol{k}} \, \xi_Y(|\boldsymbol{x}'|) \\
&= (2\pi)^3 \, \delta_\mathrm{D}(\boldsymbol{k} - \boldsymbol{k}') \, P_Y(k) \; .
\end{aligned} \tag{2.52}$$

Here, we have defined the *power spectrum* $P_Y(k)$ as the Fourier transform of the two-point correlation function.

For a Gaussian random field, the uncorrelatedness of the Fourier modes expressed by Eq. (2.52) also implies mutual statistical independence. Furthermore, the probability density for each mode is a Gaussian with a variance proportional to the power spectrum.

The power spectrum of density fluctuations

As any random field, the density contrast $\delta(\boldsymbol{x})$ can be characterized by its m-point correlation functions. Since the density contrast is defined to have vanishing mean, the first non-trivial moment is the two-point correlation function, or, equivalently, the power spectrum $P_\delta(k)$, which is defined through

$$\left\langle \tilde{\delta}(\boldsymbol{k}) \, \tilde{\delta}^*(\boldsymbol{k}') \right\rangle = (2\pi)^3 \, \delta_\mathrm{D}(\boldsymbol{k} - \boldsymbol{k}') \, P_\delta(k) \; . \tag{2.53}$$

The importance of the power spectrum in cosmology originates from the prediction of theories of Inflation that the initial density fluctuation field is a Gaussian random field. It therefore contains all information about δ at early times. A further generic property of these theories is that the initial power spectrum (i.e., at times when all density fluctuations of interest are
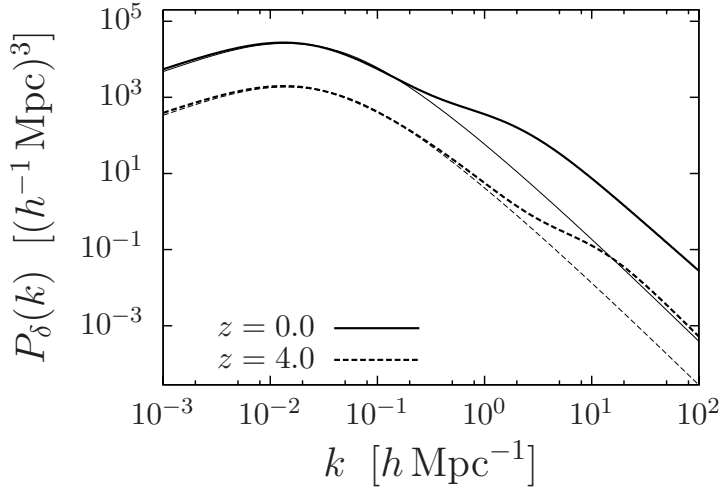
Figure 2.3: *Matter power spectra for $z = 0$ and $z = 4$, computed using the fit formula by Smith et al. (2003). Thick lines show non-linear power spectrum, whereas thin lines give P_δ in linear perturbation theory.*

outside the horizon) is given by a power law, $P_i(k; t_i) = A\, k^{n_s}$, where the *spectral index* $n_s \approx 1$ and A is a normalization constant. The power spectrum at later times can be computed using linear perturbation theory (see Sec. 2.2.2) and the transfer function defined by Eq. (2.47):

$$P_\delta(k, t) = A\, T^2(k)\, \frac{D_+^2(t)}{D_+^2(t_i)}\, k^{n_s}\,. \tag{2.54}$$

The normalization A of the power spectrum cannot be predicted from theory, but has to be measured. By convention, this is done in terms of the dispersion of the density field smoothed with a normalized, spherical top-hat filter of radius R:

$$\sigma^2(R) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\, |\tilde{W}_R(k)|^2\, P_\delta(k)\,, \tag{2.55}$$

where $\tilde{W}_R(k)$ is the Fourier-transform of the filter function. The usual choice is $R = 8\, h^{-1}$ Mpc, and the resulting dispersion is denoted by σ_8.

   The statements in the previous paragraph cease to be valid when the density contrast approaches unity and non-linear effects become important. One possibility to predict the evolution of P_δ in this case is to employ perturbation theory to compute higher-order corrections to the linear evolution equations (see e.g. Bernardeau et al. 2002). A further, more popular way to obtain a prediction for P_δ is to measure the power spectrum in N-body simulations (see Chapter 4) for various redshifts and cosmologies, and then to inter- or extrapolate the power spectrum to other cosmologies using fitting formulae. This has been done in Peacock & Dodds (1996), and more recently in Smith et al. (2003). Finally, simplified models of structure formation have been developed which allow the computation of the power spectrum and higher-order moments of the density field. In the *halo model*, it is assumed that all matter in the Universe is in the form of gravitationally bound objects, so-called *halos*. Using input from simulations, such as the mass spectrum of these halos or the halo density profile, and results from linear perturbation theory, this model is very successful in predicting properties of the density field for a large variety of cosmological models (for a review, see Cooray & Sheth 2002).

   In Fig. 2.3, we compare the matter power spectrum for $z = 0$ and $z = 4$ obtained from linear perturbation theory with the fit formula by Smith et al. (2003). The linear power spectrum grows $\propto k$ on large scales. On smaller scales, the effects described by the transfer function lead

Description	Symbol	WMAP5 only	WMAP5+BAO+SN
Age of universe [Gyr]	t_0	13.69 ± 0.13 Gyr	13.72 ± 0.12 Gyr
Hubble constant [$\mathrm{km\,s^{-1}\,Mpc^{-1}}$]	H_0	$71.9^{+2.6}_{-2.7}$	70.5 ± 1.3
Baryon density	Ω_{b}	0.0441 ± 0.0030	0.0456 ± 0.0015
Dark matter density	Ω_{dm}	0.214 ± 0.027	0.228 ± 0.013
Dark energy density	Ω_{Λ}	0.742 ± 0.030	0.726 ± 0.015
Fluctuation amplitude at $8h^{-1}$ Mpc	σ_8	0.796 ± 0.036	0.812 ± 0.026
Spectral index	n_{s}	$0.963^{+0.014}_{-0.015}$	0.960 ± 0.013
Redshift of matter-radiation equality	z_{eq}	3176^{+151}_{-150}	3253^{+89}_{-87}

Table 2.1: *Cosmological parameters for the standard model as determined from the WMAP5 CMB data alone and from the combination with baryonic acoustic oscillation data (BAO) and supernovae type Ia data (SN) (extracted from Hinshaw et al. 2008). Note that $\Omega_{\mathrm{m}} = \Omega_{\mathrm{dm}} + \Omega_{\mathrm{b}}$.*

to a decline $\propto k^{-3}$. The non-linear evolution of the density field leads to an increase of power on small scales; the length scale at which the transition between the linear and the non-linear regime occurs is larger at later times.

3 Gravitational Lensing

3.1 Introduction

Gravitational lensing is concerned with the deflection of light in the gravitational field of massive objects. Einstein's General Theory of Relativity predicts that the deflection angle, i.e. the angle between incoming and outgoing light ray, caused by a point mass M, is given by

$$\boldsymbol{\alpha} = \frac{4GM}{c^2} \frac{\boldsymbol{r}}{r^2} \, , \tag{3.1}$$

where $\boldsymbol{r}$ is the impact vector of the light ray. This causes background sources whose light passes close to a massive body (the "lens") to appear at slightly different positions and compared to the case where the lens is absent. If the source is extended, image distortions by differential deflection can be observed. Gravitational lensing has first been observed by Eddington and Dyson (1919) during a Solar eclipse, which allowed them to measure the apparent positions of stars very close to the obscured Sun and compare them to the unlensed positions recorded earlier. In a cosmological context, however, gravitational lensing was a purely theoretical matter until the discovery of the "double quasar" (Walsh et al. 1979), which proved to be a pair of images of the same source, lensed by a galaxy in the foreground. As of today, several hundreds of such multiple image systems are known. Gravitational lensing by galaxy clusters was discovered by Lynds & Petrosian (1986) in the form of strongly distorted and highly elongated images of galaxies located behind the cluster, dubbed "giant luminous arcs". Both multiple image systems as well as giant arcs are manifestations of "strong" gravitational lensing. To occur, it requires an accurate alignment of source and lens along the line of sight and/or a very massive lens.

But even in less favorable situations, lensing has its impact on the observed images of distant sources. Galaxy images at larger distances from a mass concentration will be distorted, too, although this effect cannot be detected in a single image anymore. In most cases, the distortions caused by lensing are much smaller than the *shape noise* caused by the fact that the source galaxies are not intrinsically round. This is the domain of "weak" gravitational lensing. To measure this effect requires the use of statistical methods applied to a large number of galaxies. Tyson et al. (1990) for the first time found systematic distortions of background galaxies near galaxy clusters. Weak lensing by galaxies, or *galaxy-galaxy lensing*, has been detected by Brainerd et al. (1996). Finally, *cosmic shear*, i.e. weak lensing by the large-scale structure in the Universe, was observed for the first time in Bacon et al. (2000), Kaiser et al. (2000), van Waerbeke et al. (2000) and Wittman et al. (2000).

In this chapter, we will briefly discuss the basics of gravitational lensing (mostly following the reviews of Bartelmann & Schneider 2001; Schneider 2006b), before focusing in more detail on cosmic shear and galaxy-galaxy lensing.
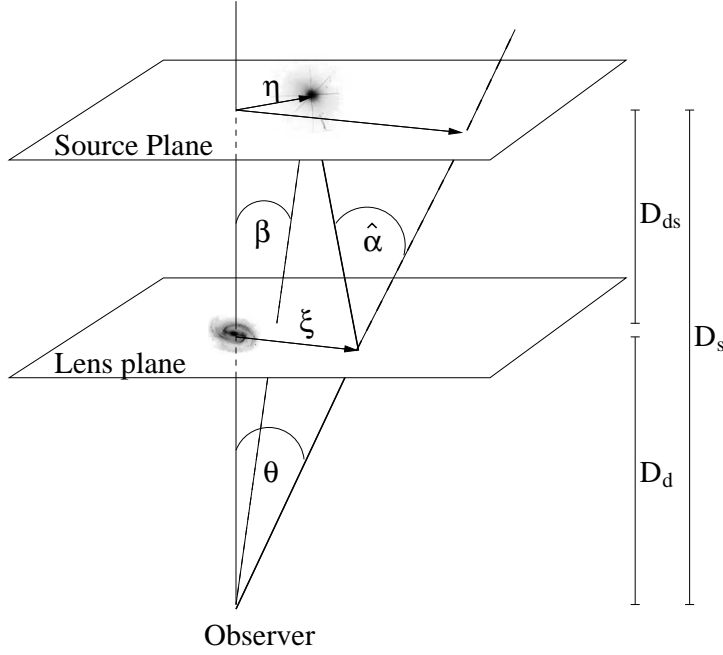
Figure 3.1: *Geometry of a typical lens system*

3.2 The thin lens approximation

3.2.1 Formalism

In most cases of strong lensing, the radial extent of the lens is much smaller than the dimensions of the whole lens system. For example, compare the size of a typical galaxy ($d \lesssim 1$ Mpc) to the typical distances between source, lens and observer, which can be of the order of 1 Gpc. Therefore, it is an excellent approximation to project the lensing mass distribution onto a plane perpendicular to the line of sight and to assume that light deflection occurs only where the ray intersects that plane. The overall setup of such a lens system is shown in Fig. 3.1. Throughout, we will denote with $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ the angular coordinates in the source and lens plane, respectively. Physical distances on the source plane are labelled with $\boldsymbol{\eta}$, distances in the lens plane with $\boldsymbol{\xi}$. The angular diameter distances to the lens plane, to the source plane and between lens and source plane are given by D_d, D_s and D_ds.

The deflection angle by the extended mass distribution on the lens plane can be computed as a superposition of point mass contributions:

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int_{\mathbb{R}^2} d^2\xi' \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{\left|\boldsymbol{\xi} - \boldsymbol{\xi}'\right|^2} \Sigma(\boldsymbol{\xi}') \ , \tag{3.2}$$

where $\Sigma(\boldsymbol{\xi}')$ is the surface mass density on the lens plane.

Assuming that the angular extent of the lens system on the sky is small, one can read off from Fig. 3.1 the lens equation

$$\boldsymbol{\eta} = \frac{D_\mathrm{s}}{D_\mathrm{d}} \boldsymbol{\xi} - D_\mathrm{ds} \hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \ . \tag{3.3}$$

Transforming to angular coordinates, so that $\boldsymbol{\eta} = D_\mathrm{s}\boldsymbol{\beta}$ and $\boldsymbol{\xi} = D_\mathrm{d}\boldsymbol{\theta}$, and introducing the scaled

deflection angle $\boldsymbol{\alpha} \equiv \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\hat{\boldsymbol{\alpha}}$, the lens equation takes the simple form

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}) \ . \tag{3.4}$$

Given a fixed source position $\boldsymbol{\beta}$, this is a nonlinear equation for the image position $\boldsymbol{\theta}$. If the lens is massive enough and lens and source are well aligned along the optical axis, Eq. (3.4) will have more than one solution and therefore yield multiple images of a single source.

Lens theory takes a very convenient form if expressed in terms of the *convergence* κ and the *lensing potential* ψ:

$$\kappa(\boldsymbol{\theta}) \equiv \frac{\Sigma(D_{\mathrm{d}}\boldsymbol{\theta})}{\Sigma_{\mathrm{crit}}} \ , \tag{3.5}$$

$$\psi(\boldsymbol{\theta}) \equiv \frac{1}{\pi} \int \mathrm{d}^2\theta' \, \kappa(\boldsymbol{\theta}') \ln|\boldsymbol{\theta} - \boldsymbol{\theta}'| \ , \tag{3.6}$$

where we have defined the critical surface mass density

$$\Sigma_{\mathrm{crit}}(z_{\mathrm{d}}, z_{\mathrm{s}}) = \frac{c^2}{4\pi G}\frac{D_{\mathrm{s}}}{D_{\mathrm{d}}D_{\mathrm{ds}}} \ . \tag{3.7}$$

Then, the scaled deflection angle is given by

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int \mathrm{d}^2\theta' \, \kappa(\boldsymbol{\theta}')\frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{\left|\boldsymbol{\theta} - \boldsymbol{\theta}'\right|^2} = \boldsymbol{\nabla}\psi(\boldsymbol{\theta}) \ . \tag{3.8}$$

Since $\boldsymbol{\nabla}^2 \ln|\boldsymbol{\theta}| = 2\pi\delta_{\mathrm{D}}(\boldsymbol{\theta})$, taking the two-dimensional Laplacian of Eq. (3.6) yields the Poisson equation

$$\boldsymbol{\nabla}^2\psi = 2\kappa \ . \tag{3.9}$$

3.2.2 The lens mapping

Eq. (3.4) defines a mapping $\boldsymbol{\theta} \to \boldsymbol{\beta}(\boldsymbol{\theta})$ from the image plane to the source plane, which in the case of multiple images is not globally invertible. If the images of the lensed object are small compared to the scales on which the deflection angle varies considerably, local information may be obtained by linearizing the lens mapping. Its Jacobian is given by

$$\mathsf{A}_{ij}(\boldsymbol{\theta}) = \frac{\partial\beta_i}{\partial\theta_j} = \delta_{ij} - \frac{\partial^2\psi(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} \ . \tag{3.10}$$

Lensing changes the solid angle subtended by the source compared to the unlensed case. Since surface brightness is conserved, the flux ratio of a lensed image and unlensed source is given by the ratio of the respective solid angles. Therefore, the magnification μ is defined as

$$\frac{1}{\mu} = \frac{S(\text{without lens})}{S(\text{with lens})} = \left|\frac{\mathrm{d}^2\beta}{\mathrm{d}^2\theta}\right| = \left|\det\mathsf{A}\right| \ . \tag{3.11}$$

If $\det\mathsf{A} = 0$, the magnification becomes formally infinite. This of course does not happen in reality: in this case finite source size and so far neglected effects of wave optics become important

(see Schneider et al. 1992). The set of points $\boldsymbol{\theta}$ for which $\det \mathsf{A} = 0$ forms closed curves, the so-called *critical curves*. Their images in the source plane are called *caustics*. If a source lies near or on such a curve, Eq. (3.11) tells that its images will be highly magnified and distorted. The giant luminous arcs found in clusters of galaxies provide a prominent example for this.

Exploiting the mapping between $\mathbb{R}^2$ and the complex numbers, any 2×2-matrix T can be written as $x \otimes y$, where $x, y \in \mathbb{C}$ and $x = x_1 + \mathrm{i} x_2$. We set $x_i y_j = \mathsf{T}_{ij}$. This allows us to identify linear combinations of the elements of T with specified transformation properties under rotations. Note that $xy \mapsto xy\,\mathrm{e}^{-2\mathrm{i}\alpha}$ and xy^* is invariant under a rotation by the angle α. Therefore, we call xy a spin-2 quantity and xy^* a spin-0 quantity. In general, a spin-n quantity z transforms as $z \mapsto z\,\mathrm{e}^{-n\mathrm{i}\alpha}$.

In terms of the matrix elements T_{ij}, we have

$$
\begin{aligned}
\text{Spin 0:} \qquad xy^* &= \ (\mathsf{T}_{11} + \mathsf{T}_{22}) + \mathrm{i}(\mathsf{T}_{21} - \mathsf{T}_{12}) \\
\text{Spin 2:} \qquad xy &= \ (\mathsf{T}_{11} - \mathsf{T}_{22}) + \mathrm{i}(\mathsf{T}_{12} + \mathsf{T}_{21})
\end{aligned}
\tag{3.12}
$$

Applying this to the case of the symmetric Jacobian of the lens mapping, we can define a spin-0 and two spin-2 quantities:

$$
\kappa \ = \ 1 - \frac{1}{2}(\mathsf{A}_{11} + \mathsf{A}_{22}) = \frac{1}{2}\left(\psi_{,11} + \psi_{,22}\right) \ , \tag{3.13}
$$

$$
\gamma_1 \ = \ \frac{1}{2}(\mathsf{A}_{22} - \mathsf{A}_{11}) = \frac{1}{2}\left(\psi_{,11} - \psi_{,22}\right) \ , \tag{3.14}
$$

$$
\gamma_2 \ = \ -\mathsf{A}_{12} = \psi_{,12} \ . \tag{3.15}
$$

$$
\tag{3.16}
$$

Here, γ_1 and γ_2 are the components of the *shear*, which are often combined to form the complex shear

$$
\gamma = \gamma_1 + \mathrm{i}\gamma_2 = |\gamma|\mathrm{e}^{2\mathrm{i}\phi} \ . \tag{3.17}
$$

The factor 2 in the exponential in Eq. (3.17) can also be understood by noting that an ellipse is transformed into itself after a rotation by 180°, not 360°. With these definitions, the Jacobian matrix can be written as

$$
\mathsf{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix} \ , \tag{3.18}
$$

where the *reduced shear* g is given by $g = \gamma/(1 - \kappa)$.

If we consider a small, intrinsically circular source centered on $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\theta}_0)$ and with intensity $I^{\mathrm{s}}(\boldsymbol{\beta})$, we can linearize the lens mapping and write for the observed intensity

$$
I(\boldsymbol{\theta}) = I^{\mathrm{s}}\left[\boldsymbol{\beta}_0 + \mathsf{A}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right] \ . \tag{3.19}
$$

We see that in the linear approximation, the image of a circular source will be isotropically (de-)magnified and stretched into a elliptical shape, the axis ratio of which is given by the ratio of the eigenvalues of A.

Since both κ and γ are second partial derivatives of the deflection potential (Eqs. 3.9, 3.14 and 3.15), it is possible to write down a relation between them in Fourier space (Kaiser-Squires relation, Kaiser & Squires 1993):

$$
\tilde{\gamma}(\boldsymbol{\ell}) = \left(\frac{\ell_\ell^2 - \ell_2^2}{|\boldsymbol{\ell}|^2} + 2\mathrm{i}\frac{\ell_1\ell_2}{|\boldsymbol{\ell}|^2}\right) \tilde{\kappa}(\boldsymbol{\ell}) = \mathrm{e}^{2\mathrm{i}\phi_\ell}\tilde{\kappa}(\boldsymbol{\ell}) \ , \tag{3.20}
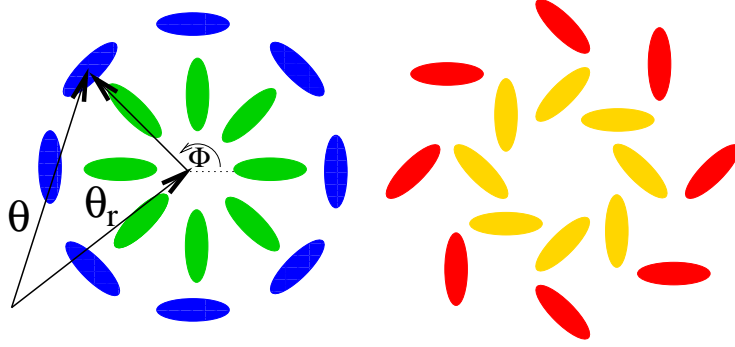$$

Figure 3.2: *Illustration of the tangential and cross components of the shear (Eq. 3.21), measured with respect to the reference point $\boldsymbol{\theta}_r$ at the center of the circles. Left panel: $\gamma_\times = 0$ and $\gamma_t = 1$ (outer ellipses), $\gamma_t = -1$ (inner ellipses). Right panel: $\gamma_t = 0$ and $\gamma_\times = -1$ (outer ellipses), $\gamma_\times = 1$ (inner ellipses).*

where $\phi_{\boldsymbol{\ell}}$ is the polar angle of $\boldsymbol{\ell}$.

For many applications, it is useful to express the shear relative to a specified direction instead of looking at the Cartesian components. One defines the tangential and cross components of the shear:

$$\gamma_t(\boldsymbol{\theta}; \boldsymbol{\theta}_r) = -\mathrm{Re}\left[\gamma(\boldsymbol{\theta})\mathrm{e}^{-2\mathrm{i}\Phi}\right]; \qquad \gamma_\times(\boldsymbol{\theta}; \boldsymbol{\theta}_r) = -\mathrm{Im}\left[\gamma(\boldsymbol{\theta})\mathrm{e}^{-2\mathrm{i}\Phi}\right], \tag{3.21}$$

where $\boldsymbol{\theta}_r$ is a reference point and Φ is the polar angle of $\boldsymbol{\theta}-\boldsymbol{\theta}_r$. As illustrated in Fig. 3.2, $\gamma_t = \pm 1$, $\gamma_\times = 0$ means a tangential or radial alignment of the shear with respect to the reference point, whereas $\gamma_\times = \pm 1$, $\gamma_t = 0$ describes curl-like patterns.

3.2.3 Ellipticity measurements

In particular in the weak lensing regime, where $\kappa \ll 1$ and $|\gamma| \ll 1$, the information about the lensing mass is hidden in the shapes of background galaxies, which of course are not intrinsically round. At present, most weak lensing analyses rely on measuring the quadrupole moment of the light distribution of each background galaxy.

The center of the brightness distribution $I(\boldsymbol{\theta})$ of a particular image is given by

$$\bar{\boldsymbol{\theta}} = \frac{\int \mathrm{d}^2\theta\, q_I[I(\boldsymbol{\theta})]\,\boldsymbol{\theta}}{\int \mathrm{d}^2\theta\, q_I[I(\boldsymbol{\theta})]}, \tag{3.22}$$

where q_I is a weight function which can account for details of how photometry and object detection are performed. The second brightness moment is defined as

$$Q_{ij} \equiv \frac{\int \mathrm{d}^2\theta\, q_I[I(\boldsymbol{\theta})](\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int \mathrm{d}^2\theta\, q_I[I(\boldsymbol{\theta})]}, \; i,j = 1,2. \tag{3.23}$$

One possibility to define the (complex) ellipticity of a galaxy is

$$\epsilon \equiv \frac{Q_{11} - Q_{22} + 2\mathrm{i}Q_{12}}{\mathrm{tr}Q + 2\sqrt{\det Q}}. \tag{3.24}$$

For a source with elliptical isophotes with axis ratio r, this yields

$$|\epsilon| = \frac{1-r}{1+r}. \tag{3.25}$$

The quadrupole tensors for unlensed source and the observed, lensed image are related by the linearized lens mapping, $\mathsf{Q}^{(s)} = \mathsf{A}\mathsf{Q}\mathsf{A}^t$. This leads to the relation between source and image ellipticities (Schneider & Seitz 1995)

$$\epsilon^{(s)} = \begin{cases} \frac{\epsilon - g}{1 - g^* \epsilon} & \text{if } |g| \leq 1 \\ \frac{1 - g\epsilon^*}{\epsilon^* - g^*} & \text{if } |g| > 1 \,. \end{cases} \tag{3.26}$$

The inverse relation is obtained by interchanging ϵ and $\epsilon^{(s)}$ and changing the sign of g. If the intrinsic ellipticities $\epsilon^{(s)}$ are oriented randomly ($\langle \epsilon^{(s)} \rangle = 0$, where $\langle \ldots \rangle$ is the expectation value of the ellipticity), one can show that

$$\langle \epsilon \rangle = \begin{cases} g & \text{if } |g| \leq 1 \\ 1/g^* & \text{if } |g| > 1 \,. \end{cases} \tag{3.27}$$

This means that the actual observable is the reduced shear (and not γ), since each image ellipticity is an unbiased estimate of g. In the weak lensing regime, however, the convergence is small, so that $g \approx \gamma$ and $\langle \epsilon \rangle = \gamma$.

3.2.4 The mass-sheet degeneracy

Observable quantities for a typical strong lens system are usually lens and image positions, image fluxes, and for some lens systems also image shapes can be measured. The aim then is to construct a lens model (i.e. a model for $\kappa(\boldsymbol{\theta})$), which reproduces the observed image properties. Unfortunately, if such a model is found, it is by no means unique, since

$$\kappa_\lambda(\boldsymbol{\theta}) = 1 + \lambda\,(\kappa(\boldsymbol{\theta}) - 1) \tag{3.28}$$

will provide an equally good fit to the data. Indeed, applying this transformation changes the lens equation to

$$\frac{\boldsymbol{\beta}}{\lambda} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{3.29}$$

which is identical to the untransformed equation except for an unobservable rescaling of the source position. The Jacobian transforms as $\mathsf{A}_\lambda = \lambda\mathsf{A}$, so that $\gamma_\lambda = \lambda\gamma$ and $1 - \kappa_\lambda = \lambda(1-\kappa)$. The last two equations show that the observable reduced shear remains unchanged. This degeneracy can only be broken if an absolute measurement of a property of the source that is affected by lensing (e.g. luminosity) is available. If the source is variable, a measurement of the time delay Δt between two different images also allows the determination of λ (provided the Hubble constant is known), since $\Delta t \mapsto \lambda\Delta t$ under the mass sheet degeneracy.

3.3 Light propagation in an inhomogeneous Universe - Cosmic Shear

So far, we have considered an isolated lens in a homogeneous background, which is a good approximation only when the lens effect of the system under consideration is sufficiently strong.

In addition to these strong lensing effects, which only occur if the alignment of source and lens is nearly perfect, every light ray coming from a distant galaxy passes through the filamentary large-scale structure and is deflected continuously. The shear resulting from the differential deflection of neighboring light rays is too weak to be detected in an individual galaxy image. However, the large-scale structure imprints a coherent distortion pattern on the image shapes of distant galaxies, which is amenable to statistical methods.

3.3.1 The Jacobian matrix

If we consider lensing by large-scale structure, deflections occur everywhere along the light path. For this reason, the single lens plane formalism introduced in the previous section cannot be applied here. In particular, in general it will not be possible to write the deflection angle $\boldsymbol{\alpha} = \boldsymbol{\theta} - \boldsymbol{\beta}$ as the gradient of a scalar potential. However, we can decompose $\boldsymbol{\alpha}$ into a gradient field and a divergence-free field (e.g. Hirata & Seljak 2003):

$$\boldsymbol{\alpha}(\boldsymbol{\theta}, w) = \boldsymbol{\nabla}\psi^{\mathrm{E}}(\boldsymbol{\theta}, w) + \boldsymbol{\nabla}_{\times}\psi^{\mathrm{B}}(\boldsymbol{\theta}, w) \ , \tag{3.30}$$

where $\boldsymbol{\nabla}_{\times} = (\partial_2, -\partial_1)$. The lens equation for a source galaxy at comoving distance w is then given by

$$\boldsymbol{\beta}(\boldsymbol{\theta}, w) = \boldsymbol{\theta} - \boldsymbol{\nabla}\psi^{\mathrm{E}}(\boldsymbol{\theta}, w) - \boldsymbol{\nabla}_{\times}\psi^{\mathrm{B}}(\boldsymbol{\theta}, w) \ , \tag{3.31}$$

and the Jacobian matrix of the lens mapping reads

$$\mathsf{A}(\boldsymbol{\theta}, w) = \begin{pmatrix} 1 - \psi^{\mathrm{E}}_{,11} - \psi^{\mathrm{B}}_{,12} & -\psi^{\mathrm{E}}_{,12} - \psi^{\mathrm{B}}_{,22} \\ -\psi^{\mathrm{E}}_{,12} + \psi^{\mathrm{B}}_{,11} & 1 - \psi^{\mathrm{E}}_{,22} + \psi^{\mathrm{B}}_{,12} \end{pmatrix} \ . \tag{3.32}$$

As in the case of a single lens plane, it is desireable to rewrite the Jacobian in terms of quantities with a physical significance. Two variants based on different decompositions of the Jacobian exist in the literature:

Polar decomposition. Any invertible matrix A can be uniquely decomposed as $\mathsf{A} = \mathsf{RS}$, where R is a rotation matrix and S a symmetric, positive-semidefinite matrix (e.g. Schneider et al. 1992). Alternatively, one can also write $\mathsf{A} = \mathsf{S}'\mathsf{R}$, with $\mathsf{S}' = \mathsf{RSR}^{-1}$. Applied to the lensing Jacobian, this means that cosmic shear manifests itself by a rotation of a galaxy image, followed by the action of a symmetric matrix (just as in the case of a single lens plane). However, the rotation is not observable because the intrinsic orientation of the source is unknown. It is therefore meaningful to define

$$\mathsf{A}(\boldsymbol{\theta}, w) = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \ . \tag{3.33}$$

Decomposition according to spin. As for the symmetric case (Eq. 3.12), we can write the Jacobian matrix of cosmic shear in terms of spin-0 and spin-2 quantities. We define

$$\kappa' \;=\; 1 - \frac{1}{2}(\mathsf{A}_{11} + \mathsf{A}_{22}) = \frac{1}{2}\left(\psi^{\mathrm{E}}_{,11} + \psi^{\mathrm{E}}_{,22}\right) \ , \tag{3.34}$$

$$\gamma'_1 \;=\; -\frac{1}{2}(\mathsf{A}_{11} - \mathsf{A}_{22}) = \frac{1}{2}\left(\psi^{\mathrm{E}}_{,11} - \psi^{\mathrm{E}}_{,22}\right) + \psi^{\mathrm{B}}_{,12} \ , \tag{3.35}$$

$$\gamma'_2 \;=\; -\frac{1}{2}(\mathsf{A}_{12} + \mathsf{A}_{21}) = \psi^{\mathrm{E}}_{,12} - \frac{1}{2}\left(\psi^{\mathrm{B}}_{,11} - \psi^{\mathrm{B}}_{,22}\right) \ , \tag{3.36}$$

$$\varpi' \;=\; -\frac{1}{2}(\mathsf{A}_{12} - \mathsf{A}_{21}) = \frac{1}{2}\left(\psi^{\mathrm{B}}_{,11} + \psi^{\mathrm{B}}_{,22}\right) \ , \tag{3.37}$$

in terms of which the Jacobian is given as

$$\mathsf{A}(\boldsymbol{\theta}, w) = \begin{pmatrix} 1 - \kappa' - \gamma'_1 & -\gamma'_2 - \varpi' \\ -\gamma'_2 + \varpi' & 1 - \kappa' + \gamma'_1 \end{pmatrix} . \tag{3.38}$$

This is analogous to thin lens theory, except for the fact that there is now the additional spin-0 quantity ϖ. From Eq. (3.30) and Eqs. (3.34) and (3.37), one finds (Hirata & Seljak 2003):

$$\kappa' = \frac{1}{2} \boldsymbol{\nabla} \cdot \boldsymbol{\alpha} , \tag{3.39}$$

$$\varpi' = \frac{1}{2} \boldsymbol{\nabla}_\times \cdot \boldsymbol{\alpha} . \tag{3.40}$$

Note that shear and convergence in Eq. (3.38) are different from those defined for the polar decomposition (Eq. 3.33). To find the relations of primed and unprimed quantities, we compute the matrix product in the polar decomposition (Eq. 3.33) and determine κ, γ and ϖ from the resulting matrix according to Eqs. (3.34) – (3.37). We obtain

$$\kappa' = 1 + (\kappa - 1) \cos \phi , \tag{3.41}$$

$$\gamma'_1 = \gamma_1 \cos \phi + \gamma_2 \sin \phi , \tag{3.42}$$

$$\gamma'_2 = \gamma_2 \cos \phi - \gamma_1 \sin \phi , \tag{3.43}$$

$$\varpi' = (\kappa - 1) \sin \phi . \tag{3.44}$$

However, we will assume throughout that the lensing effects are weak, i.e. $\kappa \ll 1$, $\gamma_i \ll 1$ and $\varpi \ll 1$. In this case, $\kappa' = \kappa$, $\gamma'_i = \gamma_i$ and $\varpi = -\phi$, as can be seen by expanding the trigonometric functions in Eqs. (3.41) – (3.44) and keeping only terms of linear order. For what follows, it will be convenient to use the form of Jacobian matrix as given in Eq. (3.38), where we from now on drop the primes on convergence and shear for simplicity of notation.

3.3.2 E- and B-modes

The Kaiser-Squires relation (Eq. 3.20) can be used to derive a useful decomposition of the shear field. If there is only a single lens plane, we expect the reconstructed convergence field κ_{KS} to be real. A non-zero imaginary part of κ_{KS} that is not compatible with noise is either a sign for uncorrected systematic effects or for the breakdown of the single-lens-plane approximation. The Fourier transforms of the real and imaginary parts of the reconstructed convergence are given by

$$\tilde{\kappa}^{\mathrm{E}}(\boldsymbol{\ell}) = \frac{1}{2} \left[\tilde{\kappa}_{\mathrm{KS}}(\boldsymbol{\ell}) + \tilde{\kappa}^*_{\mathrm{KS}}(-\boldsymbol{\ell}) \right] , \qquad \tilde{\kappa}^{\mathrm{B}}(\boldsymbol{\ell}) = \frac{1}{2\mathrm{i}} \left[\tilde{\kappa}_{\mathrm{KS}}(\boldsymbol{\ell}) - \tilde{\kappa}^*_{\mathrm{KS}}(-\boldsymbol{\ell}) \right] . \tag{3.45}$$

Using Eq. (3.20) to replace $\tilde{\kappa}^{\mathrm{KS}}$ with the shear components, we find

$$\tilde{\kappa}^{\mathrm{E}}(\boldsymbol{\ell}) = \frac{1}{2} \left[\mathrm{e}^{-2\mathrm{i}\phi_\ell} \, \tilde{\gamma}(\boldsymbol{\ell}) + \mathrm{e}^{2\mathrm{i}\phi_\ell} \, \tilde{\gamma}^*(-\boldsymbol{\ell}) \right] , \tag{3.46}$$

$$\tilde{\kappa}^{\mathrm{B}}(\boldsymbol{\ell}) = \frac{1}{2} \left[\mathrm{e}^{-2\mathrm{i}\phi_\ell} \, \tilde{\gamma}(\boldsymbol{\ell}) - \mathrm{e}^{2\mathrm{i}\phi_\ell} \, \tilde{\gamma}^*(-\boldsymbol{\ell}) \right] , \tag{3.47}$$

$$\tag{3.48}$$

where ϕ_ℓ is the polar angle of the wave vector $\boldsymbol{\ell}$. Since γ_1 and γ_2 are real functions, $\tilde{\gamma}_{1,2}(\boldsymbol{\ell}) = \tilde{\gamma}^*_{1,2}(-\boldsymbol{\ell})$, which finally leads to

$$\tilde{\kappa}^{\mathrm{E}} = \cos(2\phi_\ell) \, \tilde{\gamma}_1 + \sin(2\phi_\ell) \, \tilde{\gamma}_2 , \tag{3.49}$$

$$\tilde{\kappa}^{\mathrm{B}} = -\sin(2\phi_\ell) \, \tilde{\gamma}_1 + \cos(2\phi_\ell) \, \tilde{\gamma}_2 . \tag{3.50}$$

Writing γ in terms of $\psi^{\mathrm{E,B}}$ (Eqs. 3.35 and 3.36), we find that

$$\kappa^{\mathrm{E}} = \kappa \qquad \text{and} \qquad \kappa^{\mathrm{B}} = \varpi \ , \tag{3.51}$$

so that $\kappa^{\mathrm{KS}} = \kappa + \mathrm{i}\varpi$.

The E-/B-mode decomposition of the shear field can be obtained by again using the Kaiser-Squires relation (Eq. 3.20) to replace κ^{E} and κ^{B}:

$$\tilde{\gamma}^{\mathrm{E}}(\boldsymbol{\ell}) = \frac{1}{2}\left[\tilde{\gamma}(\boldsymbol{\ell}) + \mathrm{e}^{4\mathrm{i}\phi_\ell}\,\tilde{\gamma}^*(-\boldsymbol{\ell})\right] \ , \tag{3.52}$$

$$\tilde{\gamma}^{\mathrm{B}}(\boldsymbol{\ell}) = \frac{1}{2}\left[\tilde{\gamma}(\boldsymbol{\ell}) - \mathrm{e}^{4\mathrm{i}\phi_\ell}\,\tilde{\gamma}^*(-\boldsymbol{\ell})\right] \ . \tag{3.53}$$

3.3.3 The cosmic shear power spectrum

As noted above, weak lensing by large-scale structure can only be detected with statistical methods. The central quantity here is the power spectrum of the shear field, to which all other commonly used two-point statistics of γ are related in a relatively simple fashion. For the derivation of the E- and B-mode power spectra we mainly follow the method given in Hirata & Seljak (2003).

The E- and B-mode power spectra P_κ and P_ϖ are defined by

$$\langle \tilde{\kappa}(\boldsymbol{\ell})\,\tilde{\kappa}(\boldsymbol{\ell}')\rangle = (2\pi)^2\,\delta_{\mathrm{D}}\left(\boldsymbol{\ell}+\boldsymbol{\ell}'\right)\,P_\kappa(\ell) \ , \tag{3.54}$$

$$\langle \tilde{\varpi}(\boldsymbol{\ell})\,\tilde{\varpi}(\boldsymbol{\ell}')\rangle = (2\pi)^2\,\delta_{\mathrm{D}}\left(\boldsymbol{\ell}+\boldsymbol{\ell}'\right)\,P_\varpi(\ell) \ . \tag{3.55}$$

The starting point for the computation of these statistics is a slightly perturbed Robertson-Walker metric, which we write as

$$\mathrm{d}s^2 = a^2(\tau)\left[\left(1 + \frac{2\Phi}{c^2}\right)c^2\mathrm{d}\tau^2 - \left(1 - \frac{2\Phi}{c^2}\right)\left(\mathrm{d}w^2 + f_K^2(w)\mathrm{d}\Omega^2\right)\right] \ , \tag{3.56}$$

where w is the radial comoving distance, τ the conformal time, defined by $\mathrm{d}\tau = \mathrm{d}t/a$, $f_K(w)$ the comoving angular diameter distance and Φ the peculiar gravitational potential. If one considers an infinitesimally thin light bundle converging at the observer, then as shown in Bartelmann & Schneider (2001), the comoving separation vector $\mathbf{x}$ of a ray enclosing an angle $\boldsymbol{\theta}$ with the fiducial ray at the observer evolves as

$$\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}w^2} + K\mathbf{x} = -\frac{2}{c^2}\left[\boldsymbol{\nabla}_\perp\Phi[\mathbf{x}(\boldsymbol{\theta},w),w] - \boldsymbol{\nabla}_\perp\Phi^{(0)}(w)\right] \ . \tag{3.57}$$

Here, $\boldsymbol{\nabla}_\perp = (\partial_1,\partial_2)$ denotes the transverse comoving gradient operator, i.e. the two-dimensional gradient in the plane perpendicular to the corresponding light ray, $K = (H_0/c)^2(\Omega_{\mathrm{m}} + \Omega_\Lambda - 1)$ is the spatial curvature, and $\Phi^{(0)}(w)$ the gravitational potential along the fiducial ray. If $\boldsymbol{\nabla}\Phi \to 0$, we have the special case of light propagation in an unperturbed universe, and the above differential equation becomes homogeneous. The solution with boundary conditions $\mathbf{x}(0) = 0$ and $\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}w}\big|_{w=0} = \boldsymbol{\theta}$ is given by the angular diameter distance between the rays:

$$\mathbf{x}(\boldsymbol{\theta},w) = f_K(w)\boldsymbol{\theta} \ . \tag{3.58}$$

Since Eq. (3.57) is linear, the solution in the inhomogeneous case can be calculated by the method of Green's function, which in this case is $f_K(w)$:

$$\mathbf{x}(\boldsymbol{\theta},w) = f_K(w)\boldsymbol{\theta} - \frac{2}{c^2}\int_0^w \mathrm{d}w'\,f_K(w-w')\left[\boldsymbol{\nabla}_\perp\Phi[\mathbf{x}(\boldsymbol{\theta},w'),w'] - \boldsymbol{\nabla}_\perp\Phi^{(0)}(w')\right] \ . \tag{3.59}$$

This equation may as well be written in terms of the angular separation of the two rays:

$$\boldsymbol{\beta}(\boldsymbol{\theta}, w) = \frac{\mathbf{x}(\boldsymbol{\theta}, w)}{f_K(w)} = \boldsymbol{\theta} - \frac{2}{c^2} \int_0^w \mathrm{d}w' \, G_{w,w'} \left[\boldsymbol{\nabla}_\perp \Phi[\mathbf{x}(\boldsymbol{\theta}, w'), w'] - \boldsymbol{\nabla}_\perp \Phi^{(0)}(w') \right] \,, \quad (3.60)$$

where we have defined $G_{w,w'} = \frac{f_K(w-w')}{f_K(w)}$. Since the gravitational potential has to be computed along the perturbed ray, Eq. (3.59) is an implicit equation for the photon trajectory. A solution that is of first order in the potential can be obtained by integrating along the unperturbed light ray (*Born approximation*), which yields

$$\boldsymbol{\beta}(\boldsymbol{\theta}, w) \approx \boldsymbol{\theta} - \frac{2}{c^2} \int_0^w \mathrm{d}w' \, G_{w,w'} \left[\boldsymbol{\nabla}_\perp \Phi \left[f_K(w')\boldsymbol{\theta}, w' \right] - \boldsymbol{\nabla}_\perp \Phi^{(0)}(w') \right] \,. \quad (3.61)$$

We can derive a solution that is correct to second order in Φ by integrating Eq. (3.60) over the photon trajectory given by Eq. (3.61) and expanding the potential into a Taylor series:

$$\beta_i(\boldsymbol{\theta}, w) = \theta_i - \alpha_i^{(1)} - \alpha_i^{(2)} - \mathcal{O}(\Phi^3) \,, \qquad i \in \{1, 2\} \,, \quad (3.62)$$

where

$$\alpha_i^{(1)} = \frac{2}{c^2} \int_0^w \mathrm{d}w' \, G_{w,w'} \, \Phi_{,i} \left[f_K(w')\boldsymbol{\theta}, w' \right] \quad (3.63)$$

$$\alpha_i^{(2)} = \frac{2}{c^2} \int_0^w \mathrm{d}w' \, G_{w,w'} \, f_K(w') \, \Phi_{,ia} \left[f_K(w')\boldsymbol{\theta}, w' \right] \, \alpha_a^{(1)} [f_K(w')\boldsymbol{\theta}, w']$$

$$= -\frac{4}{c^4} \int_0^w \mathrm{d}w' \int_0^{w'} \mathrm{d}w'' \, G_{w,w'} \, G_{w',w''} \, f_K(w')$$

$$\times \, \Phi_{,ia} \left[f_K(w')\boldsymbol{\theta}, w' \right] \, \Phi_{,a} \left[f_K(w'')\boldsymbol{\theta}, w'' \right] \,. \quad (3.64)$$

Here, we use the notation $\Phi_{,i} = \partial\Phi/\partial x_i$, with $i \in \{1; 2\}$. Note that we are using the Einstein summation convention, i.e. a summation over the duplicate index a is implicit in Eq. (3.64). For notational convenience, we have dropped the terms involving the potential along the fiducial ray. Since we are only interested in derivatives of the deflection angle, they are in any case irrelevant to our calculations.

We briefly digress to derive an expression for κ in the Born approximation, which we denote with $\kappa^{(1)}$. Using Eqs. (3.39) and (3.63), we have

$$\kappa^{(1)}(\boldsymbol{\theta}, w) = \frac{1}{c^2} \int_0^w \mathrm{d}w' \, G_{w,w'} \, f_K(w') \, (\Phi_{,11} + \Phi_{,22}) \left[f_K(w')\boldsymbol{\theta}, w' \right] \,. \quad (3.65)$$

We now write $\Phi_{,11} + \Phi_{,22} = \boldsymbol{\nabla}^2 \Phi - \Phi_{,33}$. This allows us to use the Poisson equation (2.36) to replace $\boldsymbol{\nabla}^2 \Phi$ with the density contrast δ. The line-of-sight integral over $\Phi_{,33}$ is negligibly small, as has been shown in Jain et al. (2000) and White & Hu (2000). We find that to first order, κ is a projection of the three-dimensional density contrast:

$$\kappa^{(1)}(\boldsymbol{\theta}, w) = \frac{3H_0^2 \Omega_{\mathrm{m}}}{2c^2} \int_0^w \mathrm{d}w' \, \frac{G_{w,w'} \, f_K(w')}{a(w')} \, \delta \left[f_K(w')\boldsymbol{\theta}, w' \right] \,. \quad (3.66)$$

This equation applies for sources at a single redshift z. The more realistic case of source galaxies distributed in redshift according to a distribution function $p(z)$, or equivalently $p(w) =$

$p(z)|\mathrm{d}z/\mathrm{d}w|$, can be obtained as follows. The total convergence is the sum of the convergence for each source redshift, weighted by the redshift distribution:

$$
\begin{aligned}
\kappa(\boldsymbol{\theta}) &= \int \mathrm{d}w\, p(w)\, \kappa(\boldsymbol{\theta}, w) \\
&= \frac{3H_0^2 \Omega_\mathrm{m}}{2c^2} \int_0^w \mathrm{d}w'\, \frac{g(w)\, f_K(w')}{a(w')}\, \delta\left[f_K(w')\boldsymbol{\theta}, w'\right] ,
\end{aligned}
\tag{3.67}
$$

where w_h is the comoving horizon distance and

$$
g(w) = \int_w^{w_\mathrm{h}} \mathrm{d}w'\, p_w(w') \frac{f_K(w'-w)}{f_K(w')} .
\tag{3.68}
$$

Returning to the computation of the power spectra, we now make use of Eqs. (3.39) and (3.40), which in Fourier space read $\tilde{\kappa} = -(\mathrm{i}/2)\boldsymbol{\ell} \cdot \tilde{\boldsymbol{\alpha}}$ and $\tilde{\varpi} = -(\mathrm{i}/2)\,\boldsymbol{\ell} \times \tilde{\boldsymbol{\alpha}}$ (with $\boldsymbol{a} \times \boldsymbol{b} = a_1 b_2 - a_2 b_1$), to compute $\tilde{\kappa}$, $\tilde{\varpi}$ and their power spectra from the deflection angle as given by Eq. (3.62). We begin with the computation of $\tilde{\boldsymbol{\alpha}}^{(1)}$. We first carry out the two-dimensional transform of $\boldsymbol{\alpha}^{(1)}$ and then express Φ by its three-dimensional Fourier transform:

$$
\begin{aligned}
\tilde{\alpha}_i^{(1)}(\boldsymbol{\ell}, w) &= \int \mathrm{d}^2\theta\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\theta}\cdot\boldsymbol{\ell}}\, \alpha_i^{(1)}(\boldsymbol{\theta}, w) \\
&= \frac{2}{c^2} \int_0^w \mathrm{d}w'\, G_{w,w'} \int \mathrm{d}^2\theta\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\theta}\cdot\boldsymbol{\ell}}\, \Phi_{,i}\left[f_K(w')\boldsymbol{\theta}, w'\right] \\
&= \frac{2}{c^2} \int_0^w \mathrm{d}w'\, G_{w,w'} \int \mathrm{d}^2\theta\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\theta}\cdot\boldsymbol{\ell}} \int \frac{\mathrm{d}^3k}{(2\pi)^3} \\
&\quad \times \mathrm{e}^{\mathrm{i}[f_K(w')\boldsymbol{k}_\perp\cdot\boldsymbol{\theta}+k_3 f_K(w')]}\, (-\mathrm{i}\,k_i)\, \tilde{\Phi}(\boldsymbol{k}; w')
\end{aligned}
\tag{3.69}
$$

The $\boldsymbol{\theta}$-integration yields $(2\pi)^2\, \delta_\mathrm{D}\left(\boldsymbol{\ell} - f_K(w')\boldsymbol{k}_\perp\right)$. Carrying out the integral over $\boldsymbol{k}_\perp$, we set $\boldsymbol{k}_\perp = \boldsymbol{\ell}/f_K(w')$. The result is

$$
\tilde{\alpha}_i^{(1)}(\boldsymbol{\ell}, w) = -\frac{2\mathrm{i}}{c^2} \int_0^w \mathrm{d}w'\, \frac{G_{w,w'}}{f_K^3(w')} \int \frac{\mathrm{d}k_3}{2\pi} \mathrm{e}^{\mathrm{i}f_K(w')k_3}\, \ell_i\, \tilde{\Phi}\left(\frac{\boldsymbol{\ell}}{f_K(w')}, k_3; w'\right) .
\tag{3.70}
$$

To simplify the notation, we introduce the operator

$$
\mathbb{E}_w^{k_3} = \int \frac{\mathrm{d}k_3}{2\pi} \mathrm{e}^{\mathrm{i}f_K(w)k_3} .
\tag{3.71}
$$

Following the same procedure as for $\alpha_i^{(1)}$, we find for the second order correction to the deflection angle

$$
\begin{aligned}
\tilde{\alpha}_i^{(2)}(\boldsymbol{\ell}, w) &= \frac{2}{c^2} \int_0^w \mathrm{d}w'\, G_{w,w'}\, f_K(w') \int_0^{w'} \mathrm{d}w''\, G_{w',w''} \int \mathrm{d}^2\theta\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\theta}\cdot\boldsymbol{\ell}} \int \frac{\mathrm{d}^3k}{(2\pi)^3} \int \frac{\mathrm{d}^3k'}{(2\pi)^3} \\
&\quad \times \mathrm{e}^{\mathrm{i}[f_K(w')\boldsymbol{k}_\perp\cdot\boldsymbol{\theta}+k_3 f_K(w')]}\, \mathrm{e}^{\left[f_K(w'')\boldsymbol{k}_\perp'\cdot\boldsymbol{\theta}+k_3' f_K(w'')\right]} \\
&\quad \times (-k_i k_a)(-\mathrm{i}\,k_a')\, \tilde{\Phi}(\boldsymbol{k}; w')\, \tilde{\Phi}(\boldsymbol{k}'; w'') .
\end{aligned}
\tag{3.72}
$$

Now, the angular integration results in the term $\delta_\mathrm{D}\left(\boldsymbol{\ell} - f_K(w')\boldsymbol{k}_\perp - f_K(w'')\boldsymbol{k}_\perp'\right)$. We define $\boldsymbol{\lambda} = f_K(w')\boldsymbol{k}_\perp$ and $\boldsymbol{\ell}'' = \boldsymbol{\ell} - \boldsymbol{\lambda}$, and carry out the integral over $\boldsymbol{k}_\perp'$:

$$
\begin{aligned}
\tilde{\alpha}_i^{(2)}(\boldsymbol{\ell}, w) &= -\frac{4\mathrm{i}}{c^4} \int_0^w \mathrm{d}w' \int_0^{w'} \mathrm{d}w''\, \frac{G_{w,w'} G_{w',w''}}{f_K^3(w')\, f_K^3(w'')} \int \frac{\mathrm{d}^2\lambda}{(2\pi)^2}\, \mathbb{E}_{w'}^{k_3}\, \mathbb{E}_{w''}^{k_3'} \\
&\quad \times \lambda_i (\boldsymbol{\lambda} \cdot \boldsymbol{\ell}'')\, \tilde{\Phi}\left(\frac{\boldsymbol{\lambda}}{f_K(w')}, k_3; w'\right) \tilde{\Phi}\left(\frac{\boldsymbol{\ell}''}{f_K(w'')}, k_3'; w''\right) .
\end{aligned}
\tag{3.73}
$$

We can now write down the Fourier transforms of κ and ϖ:

$$\tilde{\kappa}^{(1)}(\boldsymbol{\ell}, w) = -\frac{1}{c^2} \int_0^w \mathrm{d}w' \frac{G_{w,w'}}{f_K^3(w')} \mathbb{E}_{w'}^{k_3} \ell^2 \tilde{\Phi}\left(\frac{\boldsymbol{\ell}}{f_K(w')}, k_3; w'\right) \tag{3.74}$$

$$\tilde{\kappa}^{(2)}(\boldsymbol{\ell}, w) = -\frac{2}{c^4} \int_0^w \mathrm{d}w' \int_0^{w'} \mathrm{d}w'' \frac{G_{w,w'} G_{w',w''}}{f_K^3(w') f_K^3(w'')} \int \frac{\mathrm{d}^2\lambda}{(2\pi)^2} \mathbb{E}_{w'}^{k_3} \mathbb{E}_{w''}^{k_3'}$$
$$\times (\boldsymbol{\lambda} \cdot \boldsymbol{\ell})(\boldsymbol{\lambda} \cdot \boldsymbol{\ell}'') \tilde{\Phi}\left(\frac{\boldsymbol{\lambda}}{f_K(w')}, k_3; w'\right) \tilde{\Phi}\left(\frac{\boldsymbol{\ell}''}{f_K(w'')}, k_3'; w''\right) \tag{3.75}$$

$$\tilde{\varpi}^{(1)}(\boldsymbol{\ell}, w) = 0 \tag{3.76}$$

$$\tilde{\varpi}^{(2)}(\boldsymbol{\ell}, w) = \frac{2}{c^4} \int_0^w \mathrm{d}w' \int_0^{w'} \mathrm{d}w'' \frac{G_{w,w'} G_{w',w''}}{f_K^3(w') f_K^3(w'')} \int \frac{\mathrm{d}^2\lambda}{(2\pi)^2} \mathbb{E}_{w'}^{k_3} \mathbb{E}_{w''}^{k_3'}$$
$$\times (\boldsymbol{\ell} \times \boldsymbol{\lambda})(\boldsymbol{\lambda} \cdot \boldsymbol{\ell}'') \tilde{\Phi}\left(\frac{\boldsymbol{\lambda}}{f_K(w')}, k_3; w'\right) \tilde{\Phi}\left(\frac{\boldsymbol{\ell}''}{f_K(w'')}, k_3'; w''\right) \tag{3.77}$$

We see that to the lowest (i.e. second) order in the potential, which corresponds to the Born approximation, cosmic shear does not create a B-mode.

The E-mode power spectrum in Born approximation is given by the correlator

$$\langle \tilde{\kappa}^{(1)}(\boldsymbol{\ell}) \, \tilde{\kappa}^{(1)}(\boldsymbol{\ell}') \rangle = \frac{1}{c^4} \int_0^w \mathrm{d}w' \int_0^w \mathrm{d}w'' \frac{G_{w,w'} G_{w,w''}}{f_K^3(w') f_K^3(w'')} \mathbb{E}_{w'}^{k_3} \mathbb{E}_{w''}^{k_3'}$$
$$\times \ell^2 \ell'^2 \left\langle \tilde{\Phi}\left(\frac{\boldsymbol{\ell}}{f_K(w')}, k_3; w'\right) \tilde{\Phi}\left(\frac{\boldsymbol{\ell}'}{f_K(w'')}, k_3'; w''\right) \right\rangle . \tag{3.78}$$

Now, we can use the Poisson equation (Eq. 2.36) in Fourier space

$$\tilde{\Phi}\left[\frac{\boldsymbol{\ell}}{f_K(w)}, k_3; w\right] = -\frac{3 H_0^2 \Omega_{\mathrm{m}}}{2 a(w)} \frac{1}{[\ell/f_K(w)]^2 + k_3^2} \tilde{\delta}\left[\frac{\boldsymbol{\ell}}{f_K(w)}, k_3; w\right] \tag{3.79}$$

to replace the potential with the density contrast. Since the three-dimensional matter power spectrum declines $\propto k$ on large scales, we can make the approximation that there is no correlation on scales larger than a certain coherence scale. This in turn is much smaller than the length scales for which the cosmological evolution of the density field becomes important. This means that we can set $w' \approx w''$ in the correlator in Eq. (3.78), as well as in the function G. However, we have to retain the distinction between w' and w'' for the quickly varying exponential functions in the operator $\mathbb{E}$. Eq. (3.78) now becomes

$$\langle \tilde{\kappa}^{(1)}(\boldsymbol{\ell}) \, \tilde{\kappa}^{(1)}(\boldsymbol{\ell}') \rangle = \left(\frac{3 H_0^2 \Omega_{\mathrm{m}}}{2 c^2}\right)^2 \int_0^w \mathrm{d}w' \frac{G_{w,w'}^2}{f_K^2(w') a^2(w')} \int_0^w \mathrm{d}w'' \int \frac{\mathrm{d}k_3}{2\pi} \int \frac{\mathrm{d}k_3'}{2\pi}$$
$$\times \mathrm{e}^{\mathrm{i}k_3 \, w'} \, \mathrm{e}^{\mathrm{i}k_3' \, w''} \frac{\ell^2}{[\ell/f_K(w')]^2 + k_3^2} \frac{\ell'^2}{[\ell'/f_K(w')]^2 + k_3'^2}$$
$$\times \left\langle \tilde{\delta}\left(\frac{\boldsymbol{\ell}}{f_K(w')}, k_3; w'\right) \tilde{\delta}\left(\frac{\boldsymbol{\ell}'}{f_K(w')}, k_3'; w'\right) \right\rangle . \tag{3.80}$$

The integral over w'' gives $2\pi \, \delta_{\mathrm{D}}(k_3')$, which means that only modes perpendicular to the line of sight contribute to the convergence power spectrum. If we replace the correlator of the density contrasts with the definition of the matter power spectrum, $\langle \tilde{\delta}(\boldsymbol{k}) \tilde{\delta}(\boldsymbol{k}') \rangle = (2\pi)^3 \delta_{\mathrm{D}}(\boldsymbol{k} + \boldsymbol{k}') \, P_\delta(k)$,
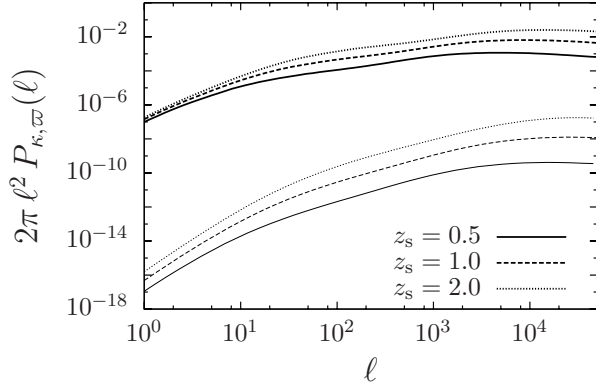
Figure 3.3: *Power spectra of the convergence κ (upper, thick curves) and the ϖ (lower, thin curves) for three different source redshifts $z_{\rm s}$. They were computed by using the three-dimensional power spectra measured from the Millennium Simulation in Eq. (3.81).*
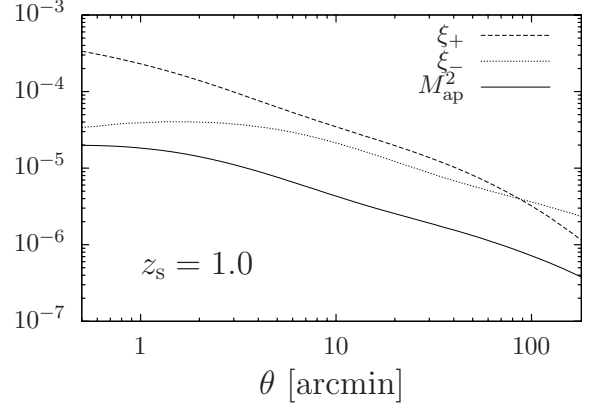
Figure 3.4: *The shear correlation functions $\xi_\pm$ and the aperture mass dispersion $M_{\rm ap}^2$ for sources at $z_{\rm s} = 1$.*

we obtain the delta functions $(2\pi)^3 f_K^2(w') \delta_{\rm D}\left(\boldsymbol{\ell} + \boldsymbol{\ell}'\right) \delta_{\rm D}\left(k_3 + k_3'\right)$. This finally enables us to carry out the integration over k_3 and k_3'. The result is

$$P_\kappa(\ell, w) = \left(\frac{3H_0^2\Omega_{\rm m}}{2c^2}\right)^2 \int_0^w \mathrm{d}w' \, \frac{G_{w,w'}^2}{a^2(w')} \, P_\delta\left(\frac{\ell}{f_K(w')}; w'\right) \, . \tag{3.81}$$

This equation applies only to source galaxies at a single redshift $z(w)$. If the sources are distributed in comoving distance, the total convergence is given by

$$\kappa(\boldsymbol{\theta}) = \int \mathrm{d}w \, p(w) \, \kappa(\boldsymbol{\theta}, w) \, . \tag{3.82}$$

Averaging Eq. (3.74) accordingly, the derivation given above stays essentially the same, except for a change in the distance weight factors. We obtain

$$P_\kappa(\ell) = \left(\frac{3H_0^2\Omega_{\rm m}}{2c^2}\right)^2 \int_0^{w_{\rm h}} \mathrm{d}w \, \frac{g^2(w)}{a^2(w)} \, P_\delta\left(\frac{\ell}{f_K(w')}; w'\right) \, , \tag{3.83}$$

where $g(w)$ is the weight function defined in Eq. (3.68).

In a very similar manner, we can compute the power spectrum of the B-mode. Since $\varpi^{(1)} = 0$, the first non-vanishing contribution to the power spectrum is given by $\langle \tilde{\varpi}^{(2)} \tilde{\varpi}^{(2)} \rangle$. This is of the order Φ^4, and therefore much smaller than the E-mode, which is of order Φ^2. We find

$$\begin{aligned} P_\varpi(\ell, w) &= 4\left(\frac{3H_0^2\Omega_{\rm m}}{2c^2}\right)^4 \int_0^w \mathrm{d}w' \, \frac{G_{w,w'}^2}{a^2(w')} \int_0^{w'} \mathrm{d}w'' \, \frac{G_{w',w''}^2}{a^2(w'')} \int \frac{\mathrm{d}^2\lambda}{(2\pi)^2} \\ &\times \frac{(\boldsymbol{\ell} \times \boldsymbol{\lambda})^2 \, (\boldsymbol{\lambda} \cdot \boldsymbol{\ell}'')^2}{\lambda^4 \, \ell''^4} \, P_\delta\left(\frac{\lambda}{f_K(w')}\right) P_\delta\left(\frac{\ell''}{f_K(w'')}\right) \, , \end{aligned} \tag{3.84}$$

where $\boldsymbol{\ell}'' = \boldsymbol{\ell} - \boldsymbol{\lambda}$.

In Fig. 3.3, we have plotted P_κ and P_ϖ for three different source redshifts. The B-mode is ≈ 6 orders of magnitude smaller than the E-mode, becoming slightly more important towards smaller scales.

Finally, note that in the derivation of Eqs. (3.81) and (3.84), the same approximations have been made that lead to Limber's equation (Limber 1953), which is used explicitly in the derivation of P_κ given in, e.g., Bartelmann & Schneider (2001) or Schneider (2006c).

3.3.4 Other two-point statistics of cosmic shear

The shear correlation functions

With the two components of the shear one can define in total four shear correlation functions. In practice, one uses γ_t and $\gamma_\times$ for a particular pair of galaxies instead of the Cartesian shear components, because the former ones are tailored to reflect the statistical isotropy of the shear field. The four possible correlation functions are $\langle \gamma_t(\boldsymbol{\theta}) \gamma_t(\boldsymbol{\theta}') \rangle$, $\langle \gamma_\times(\boldsymbol{\theta}) \gamma_\times(\boldsymbol{\theta}') \rangle$, $\langle \gamma_t(\boldsymbol{\theta}) \gamma_\times(\boldsymbol{\theta}') \rangle$ and $\langle \gamma_\times(\boldsymbol{\theta}) \gamma_t(\boldsymbol{\theta}') \rangle$, the latter two of which vanish due to parity invariance of the shear field (under a parity transformation, $\gamma_\times \to -\gamma_\times$, but $\gamma_t \to \gamma_t$). The non-vanishing correlators usually are combined to form the two shear correlation functions

$$\xi_\pm(\theta) = \langle \gamma_t(\boldsymbol{\vartheta}) \gamma_t(\boldsymbol{\theta} + \boldsymbol{\vartheta}) \rangle \pm \langle \gamma_\times(\boldsymbol{\vartheta}) \gamma_\times(\boldsymbol{\theta} + \boldsymbol{\vartheta}) \rangle \ . \tag{3.85}$$

Writing these definitions in Fourier space, $\xi_\pm$ can be related to the shear power spectrum (Kaiser 1992):

$$\xi_\pm(\theta) = \frac{1}{2\pi} \int_0^\infty \mathrm{d}\ell \, \mathrm{J}_{0,4}(\ell\theta) \, P_\kappa(\ell) \ . \tag{3.86}$$

Practical estimators for the correlation functions are given by (Schneider et al. 2002a)

$$\hat{\xi}_\pm(\theta) = \frac{1}{N_\mathrm{p}(\theta)} \sum_{ij} \left(\epsilon_{it}\epsilon_{jt} \pm \epsilon_{i\times}\epsilon_{j\times} \right) \Delta_{|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_j|, \theta} \ . \tag{3.87}$$

Here, ϵ_{it} and $\epsilon_{i\times}$ are the tangential and radial ellipticities of the i-th galaxy, which is located at $\boldsymbol{\vartheta}_i$. The symbol $\Delta_{|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_j|, \theta}$ is equal to one if $|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_j|$ falls into the angular bin centered on θ, and zero otherwise; $N_\mathrm{p}(\theta)$ is the number of pairs of galaxies that contribute to the bin considered.

The aperture mass statistic

Weak lensing observations are usually plagued by the mass sheet degeneracy, which in the regime of $\kappa \ll 1$ corresponds to adding a constant surface mass density. The aperture mass statistic M_ap (Schneider et al. 1998) was designed to be insensitive to this degeneracy. To remove any constant contribution to the convergence, κ is smoothed with a compensated filter function $U_\theta(\vartheta)$ with $\int \mathrm{d}\vartheta\, \vartheta\, U_\theta(\vartheta) = 0$:

$$M_\mathrm{ap}(\boldsymbol{\vartheta}) = \int \mathrm{d}^2\vartheta' \, \kappa(\boldsymbol{\vartheta}') \, U_\theta(|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'|) \ . \tag{3.88}$$

Here, θ is a characteristic radius of the filter function, the exact meaning of which depends on the choice of U.

The aperture mass can also be related to the tangential shear with respect to the aperture center at $\boldsymbol{\vartheta}$:

$$M_{\text{ap}}(\boldsymbol{\vartheta}) = \int \mathrm{d}^2\vartheta' \, Q_\theta(|\boldsymbol{\vartheta}'|) \, \gamma_{\text{t}}(\boldsymbol{\vartheta}';\boldsymbol{\vartheta}) \,, \tag{3.89}$$

with $Q_\theta(\vartheta) = \frac{2}{\vartheta^2} \int_0^\vartheta \mathrm{d}\vartheta' \, \vartheta' U_\theta(\vartheta') - U_\theta(\vartheta)$. A similar definition can be made for the cross component of the shear:

$$M_\perp(\boldsymbol{\vartheta}) = \int \mathrm{d}^2\vartheta' \, Q_\theta(|\boldsymbol{\vartheta}'|) \, \gamma_\times(\boldsymbol{\vartheta}';\boldsymbol{\vartheta}) \,. \tag{3.90}$$

The variances of of M_{ap} and $M_\perp$ are both related to the convergence power spectrum through a linear relation (Schneider et al. 2002b):

$$\left\langle M_{\text{ap}}^2 \right\rangle(\theta) = \frac{1}{2\pi} \int_0^\infty \mathrm{d}\ell \, \ell \, P_\kappa(\ell) \, W_{\text{ap}}^2(\theta\ell) \,, \tag{3.91}$$

$$\left\langle M_\perp^2 \right\rangle(\theta) = \frac{1}{2\pi} \int_0^\infty \mathrm{d}\ell \, \ell \, P_\varpi(\ell) \, W_{\text{ap}}^2(\theta\ell) \,, \tag{3.92}$$

where W_{ap}^2 depends on the choice of the filter function in Eq. (3.88). In general, W_{ap}^2 is a rather sharply peaked, which makes the aperture mass dispersion a good proxy of the power spectrum. Furthermore, Eq. (3.91) shows that the aperture mass dispersion has the useful property that $\left\langle M_{\text{ap}}^2 \right\rangle$ is only sensitive to the E-mode, and $\left\langle M_\perp^2 \right\rangle$ only to the B-mode.

Given a real data set, it is most straightforward to estimate the correlation functions Eq. (3.85), because they can be computed irrespectively of holes and gaps in the CCD image. Computation of M_{ap}, on the other hand, requires placing apertures on the data field, which therefore must not contain any gaps or masked regions. This is usually not possible without discarding a significant part of the CCD frame. Therefore, the best way to obtain $\left\langle M_{\text{ap}}^2 \right\rangle$ from a real data set is through the correlation functions, using (Schneider et al. 2002b):

$$\left\langle M_{\text{ap},\perp}^2 \right\rangle(\theta) = \int_0^\infty \frac{\mathrm{d}\vartheta \, \vartheta}{2\theta^2} \left[\xi_+(\vartheta) \, T_+\left(\frac{\vartheta}{\theta}\right) \pm \xi_-(\vartheta) \, T_-\left(\frac{\vartheta}{\theta}\right) \right] \,. \tag{3.93}$$

The functions T_+ and T_- depend on the choice of U in Eq. (3.88).

Popular filter functions have been proposed in Schneider et al. (2002a) and Crittenden et al. (2002). The first is a polynomial filter with finite support:

$$U_\theta(\vartheta) = \frac{9}{\pi\theta^2} \left(1 - \frac{\vartheta^2}{\theta^2}\right) \left(\frac{1}{3} - \frac{\vartheta^2}{\theta^2}\right) \mathrm{H}(\theta - \vartheta) \,. \tag{3.94}$$

For this filter, the functions relating $\left\langle M_{\text{ap}}^2 \right\rangle$ to the other two-point statistics are

$$W_{\text{ap}}^2(x) = \frac{576 \mathrm{J}_4^2(x)}{x^4}, \tag{3.95}$$

$$T_+(x) = \left\{ \frac{6(2 - 15x^2)}{5} \left[1 - \frac{2}{\pi} \arcsin\left(\frac{x}{2}\right)\right] \right. \tag{3.96}$$

$$\left. + \frac{x\sqrt{4 - x^2}}{100\pi} \left(120 + 2320x^2 - 754x^4 + 132x^6 - 9x^8\right) \right\} \mathrm{H}(2 - x), \tag{3.97}$$

$$T_-(x) = \frac{192}{35\pi} x^3 \left(1 - \frac{x^2}{4}\right)^{7/2} \mathrm{H}(2-x). \tag{3.98}$$

T_+ and T_- vanish for $x > 2$, so that $\langle M_{\mathrm{ap}}^2 \rangle$ can be obtained from the correlation functions by integration over a finite interval.

The second filter function has somewhat more convenient analytical properties, at the expense of no longer having finite support. It is given by

$$U_\theta(\vartheta) = \frac{1}{2\pi\theta^2} \left(1 - \frac{\vartheta^2}{2\theta^2}\right) \exp\left(-\frac{\vartheta^2}{2\theta^2}\right) , \tag{3.99}$$

and its auxiliary functions are

$$W_{\mathrm{ap}}^2(x) = \frac{x^4}{4} \exp\left(-x^2\right) , \tag{3.100}$$

$$T_+(x) = \left(\frac{x^4}{128} - \frac{x^2}{8} + \frac{1}{4}\right) \exp\left(-\frac{x^2}{4}\right) , \tag{3.101}$$

$$T_-(x) = \frac{x^4}{128} \exp\left(-\frac{x^2}{4}\right) . \tag{3.102}$$

To compute $\langle M_{\mathrm{ap}}^2 \rangle(\theta)$ from $\xi_\pm$ for this choice of filter function, it is in principle necessary to evaluate an integral from 0 to ∞. However, $T_\pm$ approach zero very quickly, so that in practice it is sufficient to integrate only to $\approx 3\theta$.

3.4 Galaxy-Galaxy-Lensing

In contrast to cosmic shear, galaxy-galaxy lensing (GGL) is not concerned with the large-scale distribution of matter, but with the mass associated with galaxies and the dark matter halos the galaxies reside in. Gravitational lensing is a unique tool to study these structures, because it allows probing the matter distribution directly, without additional assumptions about the dynamical state of the system under consideration.

However, a single galaxy is not massive enough to create a detectable GGL effect. For this reason, the lensing signal around many galaxies, ideally of similar type, is superposed statistically. This is done by splitting the galaxy sample into fore- and background objects, either using redshift information or simply a magnitude cut. The ellipticities of the background galaxies are expected to be tangentially aligned with respect to the lensing foreground object, and therefore usually γ_{t} and $\gamma_\times$ are estimated as functions of angular separation from the lens. This is done for each foreground-background pair, and the resulting signals are averaged, leading to a rotationally symmetric mean shear profile. This can then be compared to predictions from models of the mass distribution of and associated with the lens galaxies.

This idea can be formalized by defining the GGL signal to be the correlator of the fractional lens galaxy number density $\kappa_{\mathrm{g}}(\boldsymbol{\vartheta}_{\mathrm{L}}) = n_{\mathrm{g}}(\boldsymbol{\vartheta}_{\mathrm{L}})/\bar{n} - 1$, where n_{g} is the galaxy number density and $\bar{n}$ the mean number density on the sky, and the tangential or cross component of the shear (Hoekstra et al. 2002; Schneider 2006c):

$$\langle \gamma_{\mathrm{t},\times}(\theta) \rangle \equiv \langle \gamma_{\mathrm{t},\times}(\boldsymbol{\vartheta}_{\mathrm{S}}; \boldsymbol{\vartheta}_{\mathrm{L}}) \, \kappa_{\mathrm{g}}(\boldsymbol{\vartheta}_{\mathrm{L}}) \rangle , \tag{3.103}$$

where $\boldsymbol{\vartheta}_{\rm L}$ and $\boldsymbol{\vartheta}_{\rm S}$ are the lens and source galaxy positions, respectively, and $\theta = |\boldsymbol{\vartheta}_{\rm L} - \boldsymbol{\vartheta}_{\rm S}|$. Since the galaxy distribution is a homogeneous random field, Eq. (3.103) can also be written as

$$\langle \gamma_{{\rm t},\times}(\theta) \rangle = \langle \gamma_{{\rm t},\times}(\boldsymbol{\theta}; \mathbf{0}) \, \kappa_{\rm g}(\mathbf{0}) \rangle \ . \tag{3.104}$$

An estimator for the galaxy-galaxy lensing signal is given by

$$\hat{\gamma}_{{\rm t},\times}(\theta) = \frac{1}{N_{\rm L}} \sum_{l=1}^{N_{\rm L}} \frac{1}{N_{\rm S}(\theta; \boldsymbol{\vartheta}_l)} \sum_{s=1}^{N_{\rm S}} \epsilon_{{\rm t},\times}(\boldsymbol{\vartheta}_s; \boldsymbol{\vartheta}_l) \, \Delta_{|\boldsymbol{\vartheta}_l - \boldsymbol{\vartheta}_s|, \theta} \ . \tag{3.105}$$

Here, $\epsilon_{{\rm t},\times}(\boldsymbol{\vartheta}_s; \boldsymbol{\vartheta}_l)$ is the tangential/radial ellipticity measured from the s-th background galaxy at $\boldsymbol{\vartheta}_s$ with respect to a lens at $\boldsymbol{\vartheta}_l$. The symbol $\Delta_{|\boldsymbol{\vartheta}_l - \boldsymbol{\vartheta}_s|, \theta}$ is equal to one if $|\boldsymbol{\vartheta}_l - \boldsymbol{\vartheta}_s|$ falls into the angular bin centered on θ, and zero otherwise. Finally, $N_{\rm S}(\theta; \boldsymbol{\vartheta}_l)$ is the number of source galaxies in an annulus of radius θ which is centered on $\boldsymbol{\vartheta}_l$, and $N_{\rm L}$ is the total number of lens galaxies.

Noting that $\gamma_{\rm t} + {\rm i}\gamma_\times = -\gamma\, {\rm e}^{-2{\rm i}\Phi}$, where Φ is the polar angle of $\boldsymbol{\theta}$, we can relate the GGL signal to the statistics of the underlying galaxy and dark matter density fields:

$$\begin{aligned} \langle \gamma_{\rm t}(\theta) \rangle + {\rm i}\,\langle \gamma_\times(\theta) \rangle &= -{\rm e}^{-2{\rm i}\Phi}\, \langle \kappa_{\rm g}(\mathbf{0})\gamma(\boldsymbol{\theta}) \rangle \\ &= -\int \frac{{\rm d}^2\ell}{(2\pi)^2} \int \frac{{\rm d}^2\ell'}{(2\pi)^2}\, {\rm e}^{-{\rm i}\boldsymbol{\theta}\cdot\boldsymbol{\ell}'} {\rm e}^{-2{\rm i}\Phi}\, \langle \tilde{\kappa}_{\rm g}(\boldsymbol{\ell})\tilde{\gamma}(\boldsymbol{\ell}') \rangle \\ &= -\int \frac{{\rm d}^2\ell}{(2\pi)^2} \int \frac{{\rm d}^2\ell'}{(2\pi)^2}\, {\rm e}^{-{\rm i}\boldsymbol{\theta}\cdot\boldsymbol{\ell}'} {\rm e}^{2{\rm i}(\beta'-\Phi)}\, \langle \tilde{\kappa}_{\rm g}(\boldsymbol{\ell})\tilde{\kappa}(\boldsymbol{\ell}') \rangle \ . \end{aligned} \tag{3.106}$$

In the last step, the Kaiser-Squires relation (Eq. 3.20) was used in the form $\tilde{\gamma} = {\rm e}^{2{\rm i}\beta'}\tilde{\kappa}$, where β' is the polar angle of $\boldsymbol{\ell}'$. Note that $\alpha \equiv \beta' - \Phi$ is the angle enclosed by $\boldsymbol{\theta}$ and $\boldsymbol{\ell}'$. Introducing the cross power spectrum of the galaxy distribution and convergence, $P_{\kappa {\rm g}}(\ell) = (2\pi)^2 \delta_{\rm D}(\boldsymbol{\ell} - \boldsymbol{\ell}') \langle \tilde{\kappa}_{\rm g}(\boldsymbol{\ell})\tilde{\kappa}^*(\boldsymbol{\ell}') \rangle$, this becomes

$$\langle \gamma_{\rm t}(\theta) \rangle + {\rm i}\,\langle \gamma_\times(\theta) \rangle = -\int_0^\infty \frac{{\rm d}\ell}{(2\pi)^2}\, \ell \int_0^{2\pi} {\rm d}\alpha\, {\rm e}^{{\rm i}\ell\theta\cos\alpha} \left[\cos(2\alpha) + {\rm i}\sin(2\alpha) \right] P_{\kappa {\rm g}}(\ell) \ . \tag{3.107}$$

The imaginary part of this integral, which corresponds to $\langle \gamma_\times(\theta) \rangle$, vanishes, since the integrand consists of a product of an even and an odd function (a non-zero $\langle \gamma_\times(\theta) \rangle$ would define a preferred sense of rotation on the sky). The remaining expression gives the tangential shear around the foreground galaxies and can be cast into its final form by introducing the Bessel function J_2:

$$\langle \gamma_{\rm t}(\theta) \rangle = \frac{1}{2\pi} \int {\rm d}\ell\, \ell\, J_2(\theta\ell) P_{\kappa {\rm g}}(|\boldsymbol{\ell}|) \ . \tag{3.108}$$

Since the shear signal of many galaxies is averaged to obtain the GGL signal, it makes sense to subdivide the foreground sample into subsamples of similar galaxy type, so that the halos of the galaxies in a particular subset can be expected to have quite similar properties. For each subsample, one can now try to obtain a model for the surface mass density $\Sigma(\xi)$ associated with a lens galaxy, where ξ is the physical separation from the lens, that fits the observed $\langle \gamma_{\rm t}(\theta) \rangle$ best. The central equation for the interpretation of the GGL-signal relates the mean tangential shear on a circle of radius θ to the mean convergence $\bar{\kappa}(< \theta)$ inside the circle and the mean convergence $\langle \kappa(\theta) \rangle$ on the circle (for a derivation, see Schneider 2005):

$$\langle \gamma_{\rm t}(\theta) \rangle = \bar{\kappa}(< \theta) - \langle \kappa(\theta) \rangle \ . \tag{3.109}$$

The convergence is given by $\kappa = \Sigma/\Sigma_{\mathrm{crit}}$, where the critical surface mass density depends on the distances or redshifts of source and lens (Eq. 3.5). In the case that there is no information about the redshift of any particular galaxy, it is only possible to measure an average over the redshift distributions $p_{\mathrm{L}}(z_{\mathrm{L}})$ and $p_{\mathrm{S}}(z_{\mathrm{S}})$ of the fore- and background samples:

$$\langle \gamma_{\mathrm{t}}(\theta) \rangle = \int \mathrm{d}z_{\mathrm{L}}\, p(z_{\mathrm{L}}) \int \mathrm{d}z_{\mathrm{S}}\, p(z_{\mathrm{S}}) \frac{\bar{\Sigma}\left[< D_{\mathrm{L}}(z_{\mathrm{L}})\theta\right] - \langle \Sigma\left[D_{\mathrm{L}}(z_{\mathrm{L}})\theta\right]\rangle}{\Sigma_{\mathrm{crit}}(z_{\mathrm{L}}, z_{\mathrm{S}})}. \tag{3.110}$$

Here, we denote with D_{L} the proper angular diameter distance to the lens. Even if one assumes that all lenses have the same mass distribution, the resulting signal at fixed θ comes from light rays probing the lens galaxy's halo at different physical impact parameters $D_{\mathrm{L}}(z_{\mathrm{L}})\theta$.

If photometric or even spectroscopic redshifts are available for the foreground sample and the redshift distribution of the background galaxies is known, one can measure the GGL signal directly as a function of the physical impact parameter $\xi = D_{\mathrm{L}}\theta$:

$$\langle \gamma_{\mathrm{t}}(\xi, z_{\mathrm{L}}) \rangle = \left[\bar{\Sigma}\left(< \xi\right) - \langle \Sigma\left(\xi\right)\rangle\right] \int \mathrm{d}z_{\mathrm{S}}\, \frac{p(z_{\mathrm{S}})}{\Sigma_{\mathrm{crit}}(z_{\mathrm{L}}, z_{\mathrm{S}})} = \Delta\Sigma(\xi, z_{\mathrm{L}})\, \overline{\Sigma_{\mathrm{crit}}^{-1}}(z_{\mathrm{L}})\,, \tag{3.111}$$

where the excess surface mass density $\Delta\Sigma$ was defined. This technique has been successfully applied to SDSS data in Sheldon et al. (2004), using spectroscopic redshifts for the lenses and photometric redshift estimates for the source galaxies.

Computing $P_{\kappa g}$

In order to relate the GGL signal to the three-dimensional distributions of lens galaxies and dark matter, we have to compute the cross power spectrum $P_{\kappa g}$. First of all, we express the two-dimensional number density of galaxies, n_{g}, in terms of the three-dimensional comoving galaxy density, $n_{\mathrm{g}}^{\mathrm{3D}}$, by using the fact that lensing conserves the number of galaxies:

$$n_{\mathrm{g}}(\boldsymbol{\theta}; S) = \int_0^{w_{\mathrm{h}}} \mathrm{d}w\, \nu(w, \boldsymbol{\theta}; S)\, f_K^2(w)\, n_{\mathrm{g}}^{\mathrm{3D}}\left[\boldsymbol{x}(\boldsymbol{\theta}, w); w\right] \left|\frac{\mathrm{d}^2\beta}{\mathrm{d}^2\theta}\right|. \tag{3.112}$$

The factor $f_K^2(w)$ accounts for the fact that the field of view corresponds to a larger comoving volume at larger distances. Since we wish to go beyond the Born approximation, we distinguish between the image coordinates $\boldsymbol{\theta}$ as measured by the observer, and the source plane coordinates $\boldsymbol{\beta}$. The photon trajectory along which the projection of the galaxy distribution is performed is denoted by $\boldsymbol{x}(\boldsymbol{\theta}, w)$. The selection function ν quantifies the fraction of galaxies at comoving distance w that are included in the lens sample. For a flux-limited sample of lens galaxies with flux limit S, it is given by

$$\nu(w, \boldsymbol{\theta}; S) = \frac{n_{\mathrm{g}}^{\mathrm{3D}}\left[> S/\mu(\boldsymbol{\theta}), w\right]}{\bar{n}_{\mathrm{g}}^{\mathrm{3D}}(w)}\,, \tag{3.113}$$

where $n_{\mathrm{g}}^{\mathrm{3D}}(> S, w)$ is the three-dimensional number density of galaxies with flux larger than S at distance w and $\bar{n}_{\mathrm{g}}^{\mathrm{3D}}(w)$ is the mean galaxy number density at w. The selection function depends on the position through the magnification $\mu = 1/\det \mathsf{A}$. In the weak lensing limit, we can write $|\det \mathsf{A}| \approx 1 - 2\kappa$, and thus

$$n_{\mathrm{g}}^{\mathrm{3D}}\left[> S/\mu(\boldsymbol{\theta}), w\right] \approx n_{\mathrm{g}}^{\mathrm{3D}}(> S, w) - 2\frac{\mathrm{d}n_{\mathrm{g}}^{\mathrm{3D}}(> S, w)}{\mathrm{d}\ln S}\, \kappa(\boldsymbol{\theta}, w)\,. \tag{3.114}$$

The number density of galaxies which are above the flux limit can be expressed in terms of the observer-frame luminosity function $\Phi_L(L, w) = \mathrm{d}^2 N / \mathrm{d} L\, \mathrm{d} V$:

$$n_g^{3D}(> S, w) = \int_{D_L^2(w)\, S}^{\infty} \mathrm{d} L\ \Phi_L(L, w) \ . \tag{3.115}$$

With this, the selection function is

$$\nu(w, \boldsymbol{\theta}; S) \ = \ \nu^{(0)}(w; S) + \nu^{(1)}(w, \boldsymbol{\theta}; S) \ , \tag{3.116}$$

$$\nu^{(0)}(w; S) \ = \ \frac{n_g^{3D}(> S, w)}{\bar{n}_g^{3D}(w)} \ , \tag{3.117}$$

$$\nu^{(1)}(w, \boldsymbol{\theta}; S) \ = \ 2 \frac{D_L^2(w) S\, \Phi_L[D_L^2(w) S, w]}{\bar{n}_g^{3D}(w)}\, \kappa(\boldsymbol{\theta}, w) \ . \tag{3.118}$$

We now proceed by expressing Eq. (3.112) in terms of the 3D density contrast δ_g:

$$
\begin{aligned}
n_g(\boldsymbol{\theta}; S) = & \int_0^{w_h} \mathrm{d} w\ \nu(w, \boldsymbol{\theta}; S)\, f_K^2(w)\, |\det \mathsf{A}|\, \bar{n}_g^{3D}(w) \\
& + \int_0^{w_h} \mathrm{d} w\ \nu(w, \boldsymbol{\theta}; S)\, f_K^2(w)\, |\det \mathsf{A}|\, \bar{n}_g^{3D}(w)\, \delta_g\left[\boldsymbol{x}(\boldsymbol{\theta}, w); w\right] \ .
\end{aligned}
\tag{3.119}
$$

Introducing the mean galaxy density on the sky (using that $\langle \nu^{(1)}(w, \boldsymbol{\theta}; S) \rangle = 0$),

$$\bar{n}_g(S) = \int_0^{w_h} \mathrm{d} w\ \nu^{(0)}(w)\, f_K^2(w)\, \bar{n}_g^{3D}(w) \ , \tag{3.120}$$

the average lens galaxy distribution in comoving distance,

$$p_f(w; S) = \frac{\nu^{(0)}(w; S)\, f_K^2(w)\, \bar{n}_g^{3D}(w)}{\bar{n}_g(S)} \ , \tag{3.121}$$

and the function

$$q(w; S) = \frac{\nu^{(1)}(w, \boldsymbol{\theta}; S)\, f_K^2(w)\, \bar{n}_g^{3D}}{\bar{n}_g(S)\, \kappa(\boldsymbol{\theta}, w)} = 2 f_K^2(w) \frac{D_L^2(w) S\, \Phi_L[D_L^2(w) S, w]}{\bar{n}_g(S)} \ , \tag{3.122}$$

Eq. (3.119) can be expressed as

$$
\begin{aligned}
\kappa_g(\boldsymbol{\theta}; S) \ = & \int_0^{w_h} \mathrm{d} w\ [p_f(w; S) + q(w; S)\, \kappa(\boldsymbol{\theta}, w)]\, (|\det \mathsf{A}| - 1) + \int_0^{w_h} \mathrm{d} w\ q(w; S)\, \kappa(\boldsymbol{\theta}, w) \\
& + \int_0^{w_h} \mathrm{d} w\ [p_f(w; S) + q(w; S)\, \kappa(\boldsymbol{\theta}, w)]\ \delta_g\left[\boldsymbol{x}(\boldsymbol{\theta}, w); w\right]\, |\det \mathsf{A}| \ ,
\end{aligned}
\tag{3.123}
$$

with $\kappa_g(\boldsymbol{\theta}; S) = n_g(\boldsymbol{\theta}; S) / \bar{n}_g(S) - 1$.

Here, we have used that $\int \mathrm{d} w\, p_f(w) = 1$. If we only keep terms that are of order Φ, we obtain

$$\kappa_g^{(1)}(\boldsymbol{\theta}; S) = \int_0^{w_h} \mathrm{d} w\ p_f(w; S)\ \left[\delta_g[f_K(w)\boldsymbol{\theta}, w] - 2\kappa^{(1)}(\boldsymbol{\theta}, w)\right] + \int_0^{w_h} \mathrm{d} w\ q(w; S)\, \kappa(\boldsymbol{\theta}, w) \tag{3.124}$$

The first term of $\kappa_g^{(1)}$ is the galaxy density contrast in Born approximation; the second term accounts for the fact that the magnification effect will cause the number density on the sky to

fluctuate even for a uniform distribution of lens galaxies. This happens because in the presence of a mass overdensity, a given patch on the source plane will be stretched out to a larger area on the image plane, thus reducing the observed galaxy number density, and vice versa. The third term de- or increases the contribution to the projected density contrast from distance w depending on the sign of $\kappa(\boldsymbol{\theta}, w)$. This is due to the de- or increase of the observed flux due to the magnification effect: if $\kappa > 0$ (and therefore $\mu > 1$), intrinsically fainter galaxies can be detected than in the case of $\kappa < 0$.

The Fourier transform of $\kappa_\mathrm{g}^{(1)}$ is (dropping the dependence on S from now on)

$$
\begin{aligned}
\tilde{\kappa}_\mathrm{g}^{(1)}(\boldsymbol{\ell}) &= \tilde{\kappa}_\mathrm{g}^{(1\mathrm{A})}(\boldsymbol{\ell}) + \tilde{\kappa}_\mathrm{g}^{(1\mathrm{B})}(\boldsymbol{\ell}) + \tilde{\kappa}_\mathrm{g}^{(1\mathrm{C})}(\boldsymbol{\ell}) \\
&= \int_0^{w_\mathrm{h}} \mathrm{d}w \, p_\mathrm{f}(w; S) \left[\frac{\mathbb{E}_w^{k_3}}{f_K^2(w)} \tilde{\delta}_\mathrm{g}\left(\frac{\boldsymbol{\ell}}{f_K(w)}, k_3; w \right) \right. \\
&\qquad\qquad \left. + \frac{2}{c^2} \int_0^w \mathrm{d}w' \, \frac{G_{w,w'}}{f_K^3(w')} \mathbb{E}_{w'}^{k_3} \ell^2 \, \tilde{\Phi}\left(\frac{\boldsymbol{\ell}}{f_K(w')}, k_3; w' \right) \right] \\
&\quad - \frac{2}{c^2} \int_0^{w_\mathrm{h}} \mathrm{d}w \, q(w; S) \int_0^w \mathrm{d}w' \, \frac{G_{w,w'}}{f_K^3(w')} \mathbb{E}_{w'}^{k_3} \ell^2 \, \tilde{\Phi}\left(\frac{\boldsymbol{\ell}}{f_K(w')}, k_3; w' \right) .
\end{aligned}
\tag{3.125}
$$

The projected galaxy-dark-matter cross power spectrum for source galaxies at distance w_s is given by

$$
\langle \tilde{\kappa}_\mathrm{g}(\boldsymbol{\ell}) \, \tilde{\kappa}(\boldsymbol{\ell}', w) \rangle = (2\pi)^2 \, \delta_\mathrm{D}\left(\boldsymbol{\ell} + \boldsymbol{\ell}' \right) \, P_{\kappa \mathrm{g}}(\ell, w) .
\tag{3.126}
$$

To second order in the gravitational potential, the correlator in Eq. (3.126) consists of three terms. The first, $\langle \tilde{\kappa}_\mathrm{g}^{(1\mathrm{A})}(\boldsymbol{\ell}) \, \tilde{\kappa}^{(1)}(\boldsymbol{\ell}'; w_\mathrm{s}) \rangle$, is the cross power spectrum in Born approximation. The calculation is analogous to the one that lead to the Born level cosmic shear power spectrum (Eq. 3.81), and yields

$$
P_{\kappa \mathrm{g}}^{(1\mathrm{A})}(\ell, w_\mathrm{s}) = \frac{3 H_0^2 \Omega_\mathrm{m}}{2 c^2} \int_0^{w_\mathrm{h}} \mathrm{d}w \, \frac{p_\mathrm{f}(w; S) \, G_{w_\mathrm{s}, w}}{f_K(w) \, a(w)} \, P_{\delta \mathrm{g}}\left(\frac{\ell}{f_K(w)}; w \right) .
\tag{3.127}
$$

If the source galaxies are distributed in comoving distance, we obtain

$$
P_{\kappa \mathrm{g}}^{(1\mathrm{A})}(\ell) = \frac{3 H_0^2 \Omega_\mathrm{m}}{2 c^2} \int_0^{w_\mathrm{h}} \mathrm{d}w \, \frac{p_\mathrm{f}(w; S) \, g(w)}{f_K(w) \, a(w)} \, P_{\delta \mathrm{g}}\left(\frac{\ell}{f_K(w)}; w \right) ,
\tag{3.128}
$$

where $g(w)$ was defined in Eq. (3.68). The second term of order Φ^2 that contributes to $P_{\kappa \mathrm{g}}$ is $\langle \tilde{\kappa}_\mathrm{g}^{(1\mathrm{B})}(\boldsymbol{\ell}) \, \tilde{\kappa}^{(1)}(\boldsymbol{\ell}'; w_\mathrm{s}) \rangle$, resulting in

$$
P_{\kappa \mathrm{g}}^{(1\mathrm{B})}(\ell, w_\mathrm{s}) = -2 \left(\frac{3 H_0^2 \Omega_\mathrm{m}}{2 c^2} \right)^2 \int_0^{w_\mathrm{h}} \mathrm{d}w \, p_\mathrm{f}(w; S) \int_0^w \mathrm{d}w' \, \frac{G_{w_\mathrm{s}, w'} \, G_{w, w'}}{a^2(w')} \, P_\delta\left(\frac{\ell}{f_K(w')}; w' \right) .
\tag{3.129}
$$

Finally, for $\langle \tilde{\kappa}_\mathrm{g}^{(1\mathrm{C})}(\boldsymbol{\ell}) \, \tilde{\kappa}^{(1)}(\boldsymbol{\ell}'; w_\mathrm{s}) \rangle$, we find

$$
P_{\kappa \mathrm{g}}^{(1\mathrm{C})}(\ell, w_\mathrm{s}) = 2 \left(\frac{3 H_0^2 \Omega_\mathrm{m}}{2 c^2} \right)^2 \int_0^{w_\mathrm{h}} \mathrm{d}w \, q(w; S) \int_0^w \mathrm{d}w' \, \frac{G_{w_\mathrm{s}, w'} \, G_{w, w'}}{a^2(w')} \, P_\delta\left(\frac{\ell}{f_K(w')}; w' \right) .
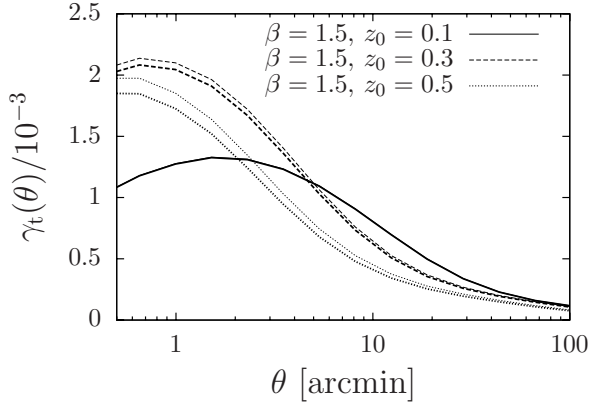\tag{3.130}
$$

Figure 3.5: *Galaxy-Galaxy lensing signal (Eq. 3.108) for three different lens redshift distributions. Shown are the Born level predictions (thin lines) and the signal including the second order corrections from Eqs. (3.128) and (3.129) (thick lines). All source galaxies are at $z = 2$.*
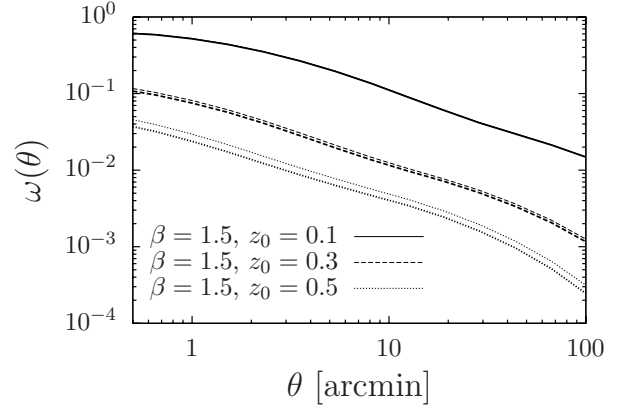
Figure 3.6: *Angular correlation functions for three different redshift distributions. Shown are the Born level predictions (thin lines) and the signal including the second order corrections from Eqs. (3.137) and (3.139) (thick lines).*

All further corrections are of fourth and higher order in the gravitational potential and therefore much smaller than the two terms considered here. For the case of a volume-limited survey, i.e. in the absence of a flux limit, Fig. 3.5 illustrates the importance of the correction term $P_{\kappa g}^{(1B)}$ for different lens redshift distributions. For low redshifts, the magnification effect is too weak to produce any noticeable deviation from the Born approximation. However, for mean redshifts beyond $z \approx 0.7$, the Born level prediction is accurate only to $\approx 7\%$.

3.5 Effects of gravitational lensing on the angular clustering power spectrum

In this section, we will derive lensing corrections to the angular clustering of galaxies, described by the angular correlation function $\omega(\theta) = \langle \kappa_g(\boldsymbol{\vartheta}) \kappa_g(\boldsymbol{\vartheta} + \boldsymbol{\theta}) \rangle$. This has already been investigated by, e.g., Villumsen (1995), Moessner et al. (1998) and Loverde et al. (2008); here, we give an alternative derivation based on our previous results for the corrections to the galaxy-galaxy lensing signal.

Since the correlation function is the Fourier transform of the angular power spectrum P_ω, the following relation holds:

$$\omega(\theta) = \frac{1}{2\pi} \int d\ell \, \ell \, J_0(\theta\ell) \, P_\omega(\ell) \, . \tag{3.131}$$

With Eq. (3.125), it is now straightforward to compute P_ω and the lowest order lensing corrections. To second order in Φ, the power spectrum is given by

$$\langle \tilde{\kappa}_g^{(1)}(\boldsymbol{\ell}) \tilde{\kappa}_g^{(1)}(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta_D \left(\boldsymbol{\ell} - \boldsymbol{\ell}' \right) P_\omega(\ell) \, , \tag{3.132}$$

where

$$\begin{aligned}
\left\langle \tilde{\kappa}_g^{(1)} \tilde{\kappa}_g^{(1)} \right\rangle &= \left\langle \tilde{\kappa}_g^{(1A)} \tilde{\kappa}_g^{(1A)} \right\rangle + \left\langle \tilde{\kappa}_g^{(1B)} \tilde{\kappa}_g^{(1B)} \right\rangle + \left\langle \tilde{\kappa}_g^{(1C)} \tilde{\kappa}_g^{(1C)} \right\rangle \\
&\quad + 2 \left\langle \tilde{\kappa}_g^{(1A)} \tilde{\kappa}_g^{(1B)} \right\rangle + 2 \left\langle \tilde{\kappa}_g^{(1A)} \tilde{\kappa}_g^{(1C)} \right\rangle + 2 \left\langle \tilde{\kappa}_g^{(1B)} \tilde{\kappa}_g^{(1C)} \right\rangle \, ,
\end{aligned} \tag{3.133}$$

and hence

$$P_\omega^{(1)}(\ell) = P_\omega^{(1A)}(\ell) + P_\omega^{(1B)}(\ell) + P_\omega^{(1C)}(\ell) + P_\omega^{(1AB)}(\ell) + P_\omega^{(1AC)}(\ell) + P_\omega^{(1BC)}(\ell) \, . \quad (3.134)$$

The first correlator in Eq. (3.133) yields the angular power spectrum in the Born approximation:

$$P_\omega^{(1A)}(\ell) = \int_0^{w_h} \mathrm{d}w \, \frac{p_f^2(w)}{f_K^2(w)} P_g\left(\frac{\ell}{f_K(w)}; w\right) \, , \quad (3.135)$$

where P_g is the three-dimensional power spectrum of the galaxy density contrast δ_g.

For $P_\omega^{(1AB)}$, we obtain

$$
\begin{aligned}
\left\langle \tilde{\kappa}_g^{(1A)}(\boldsymbol{\ell}) \tilde{\kappa}_g^{(1B)}(\boldsymbol{\ell}') \right\rangle &= \frac{2}{c^2} \int_0^{w_h} \mathrm{d}w \int_0^{w_h} \mathrm{d}\chi \int_0^\chi \mathrm{d}\chi' \, \frac{p_f(w)}{f_K^2(w)} p_f(\chi) \frac{G_{\chi,\chi'}}{f_K^3(\chi')} \\
&\quad \times \mathbb{E}_w^{k_3} \mathbb{E}_{\chi'}^{k_3'} \ell^2 \left\langle \tilde{\delta}_g\left(\frac{\boldsymbol{\ell}}{f_K(w)}, k_3; w\right) \tilde{\Phi}\left(\frac{\boldsymbol{\ell}'}{f_K(\chi')}, k_3'; \chi'\right) \right\rangle \\
&= -\frac{3H_0^2 \Omega_m}{c^2} \int_0^{w_h} \mathrm{d}w \int_0^{w_h} \mathrm{d}\chi' \int_{\chi'}^{w_h} \mathrm{d}\chi \, \frac{p_f(w)}{f_K^2(w)} \frac{p_f(\chi) \, G_{\chi,\chi'}}{f_K(\chi') \, a(\chi')} \\
&\quad \times \mathbb{E}_w^{k_3} \mathbb{E}_{\chi'}^{k_3'} \left\langle \tilde{\delta}_g\left(\frac{\boldsymbol{\ell}}{f_K(w)}, k_3; w\right) \tilde{\delta}\left(\frac{\boldsymbol{\ell}'}{f_K(\chi')}, k_3'; \chi'\right) \right\rangle \, , \quad (3.136)
\end{aligned}
$$

where in the last step we have used the Poisson equation (Eq. 2.36) and interchanged the order of the integrals over χ and χ'. Again assuming that the correlator $\langle \delta_g \, \delta \rangle$ is different from zero only if $w \approx \chi'$, we can replace χ' with w everywhere (including the lower limit of the χ-integral), except in $\mathbb{E}_{\chi'}^{k_3'}$. From here, we can proceed as before to obtain the final result

$$P_\omega^{(1AB)}(\ell) = -\frac{6H_0^2 \Omega_m}{c^2} \int_0^{w_h} \mathrm{d}w \, \frac{p_f(w) \, g_f(w)}{f_K(w) \, a(w)} \, P_{\delta g}\left(\frac{\ell}{f_K(w)}; w\right) \, , \quad (3.137)$$

where we have defined in analogy to Eq. (3.68)

$$g_f(w) = \int_w^{w_h} \mathrm{d}\chi \, p_f(\chi) \, G_{\chi,w} \, . \quad (3.138)$$

In the same manner, the third contribution to $P_\omega^{(1)}$ can be shown to be

$$P_\omega^{(1B)}(\ell) = 4\left(\frac{3H_0^2 \Omega_m}{2c^2}\right)^2 \int_0^{w_h} \mathrm{d}w \, \frac{g_f^2(w)}{a^2(w)} \, P_\delta\left(\frac{\ell}{f_K(w)}; w\right) \, . \quad (3.139)$$

Note that $\tilde{\kappa}_g^{(1C)}(\boldsymbol{\ell})$ is of the same form as $\tilde{\kappa}_g^{(1B)}(\boldsymbol{\ell})$ (see Eq. 3.125). The power spectrum corrections $P_\omega^{(1AC)}$, $P_\omega^{(1BC)}$ and $P_\omega^{(1CC)}$ can therefore be obtained from $P_\omega^{(1AB)}$ and $P_\omega^{(1B)}$ by replacing $p_f(w)$ with $-q(w; S)$ accordingly. Defining

$$h(w; S) = \int_w^{w_h} \mathrm{d}\chi \, q(\chi; S) \, G_{\chi,w} \, , \quad (3.140)$$

we find

$$
\begin{aligned}
P_\omega^{(1AC)}(\ell) &= \frac{6H_0^2 \Omega_m}{c^2} \int_0^{w_h} \mathrm{d}w \, \frac{p_f(w) \, h(w; S)}{f_K(w) \, a(w)} \, P_{\delta g}\left(\frac{\ell}{f_K(w)}; w\right) \, , \quad &(3.141)
\end{aligned}
$$

$$
\begin{aligned}
P_\omega^{(1BC)}(\ell) &= -4\left(\frac{3H_0^2 \Omega_m}{2c^2}\right)^2 \int_0^{w_h} \mathrm{d}w \, \frac{g_f(w) \, h(w; S)}{a^2(w)} \, P_\delta\left(\frac{\ell}{f_K(w)}; w\right) \, , \quad &(3.142)
\end{aligned}
$$

$$
\begin{aligned}
P_\omega^{(1C)}(\ell) &= 4\left(\frac{3H_0^2 \Omega_m}{2c^2}\right)^2 \int_0^{w_h} \mathrm{d}w \, \frac{h^2(w; S)}{a^2(w)} \, P_\delta\left(\frac{\ell}{f_K(w)}; w\right) \, . \quad &(3.143)
\end{aligned}
$$

The correction terms $P_\omega^{(1B)}$ and $P_\omega^{(1AB)}$ to the Born-level clustering power spectrum can be interpreted as follows: $P_\omega^{(1B)}$ is present even if $\kappa_g = 0$, and accounts for the correlation induced by the magnification effect caused by the large scale structure in the foreground. If $\kappa_g \neq 0$, $P_\omega^{(1AB)}$ reduces the power spectrum and therefore the angular correlation of the lens galaxies. This can be understood by considering a concentration of galaxies at distance w_g. If the galaxy distribution is correlated with the dark matter distribution ($P_{\delta g} \neq 0$), there will also be an overdensity of dark matter at this location. The magnification effect caused by this galaxy cluster will then reduce the observed number density of galaxies which are behind the cluster. This also reduces the total observed number density at this position compared to the unlensed case. Thus, as was the case for the GGL signal, the magnification bias causes an anti-correlation between galaxy populations at different distances. This effect is absent if all lens galaxies are at the same distance w_g, because then $g_f(w_g) = 0$ in Eq. (3.137).

The importance of the lensing corrections for a volume limited sample (i.e., ignoring the corrections that involve the luminosity function) can be assessed from Fig. 3.6. The Born approximation works well below mean redshifts of $\bar{z} \approx 0.3$. Beyond this redshift, P_ω^{1AB} (Eq. 3.137) becomes increasingly important, resulting in a negative correction of $\approx 10\%$ at $\bar{z} \approx 0.7$. Compared to this, the contribution from P_ω^{1B} (Eq. 3.137) is completely neglegible.

4 Simulations of structure formation

We have seen in Sec. 2.2 how a description of the evolution of the matter distribution in the Universe can be achieved in the linear regime ($|\delta| \ll 1$). The full non-linear case, however, is considerably more difficult. The evolution of the phase-space distribution of the dark matter $f(\boldsymbol{r}, \boldsymbol{u}, t)$ is described by the Vlasov equation

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \boldsymbol{u}\frac{\partial f}{\partial \boldsymbol{r}} - \boldsymbol{\nabla}\Phi\frac{\partial f}{\partial \boldsymbol{u}} = 0 \ , \tag{4.1}$$

supplemented by the Poisson equation for the gravitational field:

$$\nabla^2\Phi = 4\pi G\,\rho(\boldsymbol{r}, t) - \Lambda \ , \ \text{where} \ \rho(\boldsymbol{r}, t) = m_{\mathrm{DM}}\int \mathrm{d}^3u\,f(\boldsymbol{r}, \boldsymbol{u}, t) \ . \tag{4.2}$$

No analytical solution to these equations is known in a cosmological context, and even a numerical solution is very difficult to obtain. The simple approach of discretizing phase space and solving the system on a six-dimensional grid is doomed to fail because of the considerable computational requirements. For instance, the memory consumption of such a grid code is proportional to N_{g}^6, which means that already for $N_{\mathrm{g}} = 64$ grid cells per dimension, storing the phase space density f would require 256 GB of memory. This makes simulations of this kind of a cosmological volume with reasonable small-scale resolution virtually impossible.

Fortunately, it is not necessary to maintain the same numerical resolution everywhere in the simulation, as would be the case when using a single grid. Gravitational clustering will cause matter to concentrate in rather small volumes, leaving large, almost featureless regions nearly devoid of matter. It is therefore sufficient to concentrate most of the available computational resources on high-density regions. This can be exploited by a Monte-Carlo integration of the Vlasov-Poisson system. At early times, when the density in the simulation volume still is very smooth, the phase space is divided into a number of cells, and the mass in each cell is assigned to a pseudo-particle. These particles are evolved in time according to their equations of motion. The phase-space density f can then be estimated at any desired time from the distribution of the pseudo-particles. Since regions of high density are sampled by more particles, the resolution of the simulation automatically adapts to the matter distribution in the simulation volume. All N-body simulations used in this thesis are particle-based, which is why in this section we focus exclusively on this approach.

The simplest algorithm to evolve the particle distribution in time is to compute the gravitational force for every pair of particles, followed by a time integration step. However, this brute-force method very quickly becomes prohibitively expensive as the number of particles N grows, since the amount of operations for each time-step scales as N^2. In the following, we discuss more sophisticated algorithms that allow simulations with a much larger number of particles to be carried out than it would be possible by direct computation of the forces.

4.1 The Particle-Mesh algorithm

The particle-mesh (PM) algorithm (e.g. Hockney & Eastwood 1988) overcomes the N^2-scaling of the force computation by assigning the particles to a grid and solving the Poisson equation

using the *Fast Fourier Transform* (FFT, Cooley & Tukey 1965; Frigo & Johnson 1998). The operations count for the potential computation in this method is $\propto N_{\mathrm{g}}^3 \log N_{\mathrm{g}}^3$, where N_{g} is the number of grid cells in each direction.

4.1.1 Setting up the initial conditions

The density contrast at sufficiently early times is well described by linear perturbation theory (Sec. 2.2.2). For this case, analytical predictions for the power spectrum are readily available and can be used to generate initial conditions for the N-body simulation. To obtain the initial density and velocity field, given the power spectrum from linear theory, one makes use of the *Zel'Dovich-approximation* (Zel'Dovich 1970), which is a generalization of the concept of comoving coordinates: in a perfectly homogeneous universe, the comoving density is constant and all information about the evolution of $\rho(\boldsymbol{r})$ is contained in the mapping from comoving to proper coordinates, $\boldsymbol{r}(t) = a(t)\boldsymbol{x}(t)$. In the presence of small density fluctuations, one can make use of the fact that in the linear regime the density contrast is a separable function of $\boldsymbol{x}$ and t (Eq. 2.40). Therefore, one can relate Eulerian comoving coordinates $\boldsymbol{x}$ to Lagrangian coordinates $\boldsymbol{q}$ by the ansatz

$$\boldsymbol{x}(t) = \boldsymbol{q} - b(t)\boldsymbol{f}(\boldsymbol{q}) \ . \tag{4.3}$$

Because of the conservation of mass, the densities in Lagrangian and Eulerian coordinates are related through the determinant of the Jacobian matrix of the mapping $\boldsymbol{q}(\boldsymbol{x})$:

$$\begin{aligned}
\rho(\boldsymbol{x}) &= \rho_0 \det \left(\mathsf{E} - b(t)\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{q}} \right)^{-1} \\
&= \rho_0 \left\{ [1 - b(t)\alpha(\boldsymbol{q})][1 - b(t)\beta(\boldsymbol{q})][1 - b(t)\gamma(\boldsymbol{q})] \right\}^{-1} \\
&\approx \rho_0 \left\{ 1 - b(t)[\alpha(\boldsymbol{q}) + \beta(\boldsymbol{q}) + \gamma(\boldsymbol{q})] \right\} \ ,
\end{aligned} \tag{4.4}$$

where ρ_0 is the mean density in the Universe, and α, β and γ are the eigenvalues of the Jacobian (the Jacobian is symmetric and hence can be diagonalized provided $\boldsymbol{f}$ is irrotational). In the last step we have assumed that the eigenvalues are small. We can rewrite Eq. (4.4) in terms of the density contrast:

$$\delta(\boldsymbol{x}) = -b(t)\left[\alpha(\boldsymbol{q}) + \beta(\boldsymbol{q}) + \gamma(\boldsymbol{q})\right] \ , \tag{4.5}$$

which by comparison with Eq. (2.40) allows the identification of $b(t)$ with the negative of the growth factor, $b(t) = -D_+(t)$, since this is the only time-dependent term on the right-hand side. In the basis which diagonalizes the Jacobian, $\alpha(\boldsymbol{q}) + \beta(\boldsymbol{q}) + \gamma(\boldsymbol{q})$ is the same as the divergence of $\boldsymbol{f}$. Since the divergence is invariant under rotations, Eq. (4.4) can be rewritten as

$$\delta(\boldsymbol{x}, t) = D_+(t)\, \boldsymbol{\nabla}_{\boldsymbol{q}} \boldsymbol{f}(\boldsymbol{q}) \ . \tag{4.6}$$

The Fourier transform of $\delta(\boldsymbol{x}, t)$ is

$$\begin{aligned}
\tilde{\delta}(\boldsymbol{k}, t) &= \int \mathrm{d}^3 q \, \exp\left\{ -\mathrm{i}\boldsymbol{k} \cdot [\boldsymbol{q} + D_+(t)\boldsymbol{f}(\boldsymbol{q})] \right\} D_+(t)\, \boldsymbol{\nabla}_{\boldsymbol{q}} \boldsymbol{f}(\boldsymbol{q}) \\
&\approx \int \mathrm{d}^3 q \, \exp\left(-\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{q} \right) D_+(t)\, \boldsymbol{\nabla}_{\boldsymbol{q}} \boldsymbol{f}(\boldsymbol{q}) \\
&= -\mathrm{i}D_+(t)\, \boldsymbol{k} \cdot \tilde{\boldsymbol{f}}(\boldsymbol{k}) \ ,
\end{aligned} \tag{4.7}$$

where we have neglected all terms of higher than linear order in the perturbation $\boldsymbol{f}$. Multiplying by $\boldsymbol{k}$, we finally find

$$\tilde{\boldsymbol{f}}(\boldsymbol{k}) = \frac{\mathrm{i}\boldsymbol{k}}{|\boldsymbol{k}|^2} \frac{\tilde{\delta}(\boldsymbol{k}, t)}{D_+(t)} \ . \tag{4.8}$$

With this, it is possible to create a realization of each component of $\boldsymbol{f_k}$ in Fourier space. This is done by drawing the real and imaginary parts of $\tilde{\delta}_{\boldsymbol{k}}$ from a Gaussian random number generator with dispersion $\sqrt{L^3 P_\delta(k)/2}$, where $P_\delta(k)$ is the desired power spectrum of the density field and L is the side length of the simulation volume. Transforming back to real space, one obtains $\boldsymbol{f}(\boldsymbol{q})$. This can then be applied to the particles using Eq. (4.3), which before that were distributed uniformly in the simulation volume. This yields not only the particle positions in Eulerian space, but by differentiation with respect to t also the initial velocities.

The initial particle placement deserves some attention in order to avoid numerical artifacts. The simplest method is to start the particles from a uniform grid, but this is believed to create unrealistic structures at later times (see e.g. Smith et al. 2003). A uniform random distribution, on the other hand, will give rise to shot noise due to the Poisson sampling of the continuous density fields. A better method is to start from a *glass distribution* (White 1993), in which a random particle distribution is evolved with sign-reversed gravity into a sub-random distribution.

4.1.2 Solving the Poisson equation

At each time step, the particles are assigned to a $N_{\mathrm{g}} \times N_{\mathrm{g}} \times N_{\mathrm{g}}$ grid. This is done by giving each particle a shape $S(\boldsymbol{x})$ and assigning to each grid cell the fraction of mass that falls into the cell (Hockney & Eastwood 1988):

$$\rho(\boldsymbol{x_a}) = \frac{m_{\mathrm{p}}}{\Delta^3} \sum_{n=1}^{N} W(\boldsymbol{x}_n - \boldsymbol{x_a}) \ , \tag{4.9}$$

where

$$W(\boldsymbol{x}_n - \boldsymbol{x_a}) = \int_{x_i - \Delta/2}^{x_i + \Delta/2} \mathrm{d}x \int_{y_j - \Delta/2}^{y_j + \Delta/2} \mathrm{d}y \int_{z_k - \Delta/2}^{z_k + \Delta/2} \mathrm{d}z \, S(\boldsymbol{x} - \boldsymbol{x}_n) \ . \tag{4.10}$$

Here, Δ is the grid spacing, m_{p} is the particle mass, $\boldsymbol{x}_n$ is the position of the n-th particle and $\boldsymbol{x_a} = (x_i, y_j, z_k)$ is the location of the grid point with indices $\boldsymbol{a} = (i, j, k)$. Common particle shapes are the *nearest-grid-point* (NGP) scheme, where $S(\boldsymbol{x} - \boldsymbol{x}_n) \propto \delta_{\mathrm{D}}(\boldsymbol{x} - \boldsymbol{x}_n)$, the *cloud-in-cell* (CIC) scheme with $S(\boldsymbol{x} - \boldsymbol{x}_n) \propto \Pi(\boldsymbol{x} - \boldsymbol{x}_n)$ and the *triangular-shaped-cloud* (TSC) scheme, for which $S(\boldsymbol{x} - \boldsymbol{x}_n) \propto (\Pi * \Pi)(\boldsymbol{x} - \boldsymbol{x}_n)$. Here, we have used the rectangle function

$$\Pi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } -\Delta/2 \le \{x, y, z\} < \Delta/2 \\ 0 & \text{otherwise} \ . \end{cases} \tag{4.11}$$

Higher order schemes yield smoother density and force estimates, but become increasingly costly as more grid points are involved. Most commonly used is the CIC scheme, which provides a good compromise between smoothness and computational efficiency.

Transforming Eq. (4.2) to Fourier space, we can write the potential as

$$\tilde{\Phi}_{\boldsymbol{k}} = \tilde{G}(\boldsymbol{k}) \, \tilde{\delta}_{\boldsymbol{k}} \ , \tag{4.12}$$

where $\tilde{G}$ is the Fourier transform of Green's function. In the case of a continuous density and potential, $G(\boldsymbol{k}) \propto |\boldsymbol{k}|^{-2}$. On a grid, however, $\tilde{G}$ depends on the discretization scheme adopted. If one uses the symmetric seven-point approximation to $\boldsymbol{\nabla}^2$ (e.g. Abramowitz & Stegun 1964), it reads

$$\tilde{G}(\boldsymbol{k}) \propto \frac{1}{\sin^2(k_x\,\Delta/2) + \sin^2(k_y\,\Delta/2) + \sin^2(k_z\,\Delta/2)}\ . \tag{4.13}$$

Thus, the computation of Φ requires two Fourier transforms: $\delta \to \tilde{\delta}$ and $\tilde{\Phi} \to \Phi$.

The gravitational forces at the grid points are obtained by finite differencing of the potential. Finally, the forces are interpolated back to the particle positions. In order to avoid particle self-forces (i.e. to conserve momentum), it is important to choose the same interpolation scheme as used for the assignment of the particles to the grid.

4.1.3 Time integration

The equations of motion of the particles in the simulation can be found by requiring the conservation of the phase-space density along the trajectory $\boldsymbol{\tau}(s) = [\boldsymbol{r}(s), \boldsymbol{u}(s)]^{\mathrm{t}}$ of a particle:

$$\frac{\mathrm{d}f[\boldsymbol{\tau}(s)]}{\mathrm{d}s} = \frac{\partial f}{\partial \boldsymbol{r}}\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s} + \frac{\partial f}{\partial \boldsymbol{u}}\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}s} + \frac{\partial f}{\partial t}\frac{\mathrm{d}t}{\mathrm{d}s} = 0\ . \tag{4.14}$$

Comparison to the Vlasov equation (Eq. 4.1) then yields

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s} = \boldsymbol{u}\ ,\quad \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}s} = -\boldsymbol{\nabla}\Phi\ ,\quad \frac{\mathrm{d}t}{\mathrm{d}s} = 1\ . \tag{4.15}$$

With the initial conditions and the gravitational potential computed as described in the previous section, these equations can now be solved in discrete time-steps. A common time-stepping scheme is the *leapfrog* scheme, which is accurate to second order in the time step. This integration scheme proves to be superior to other schemes of the same order, such as the Runge-Kutta integrator (Springel 2005). It therefore represents a good compromise between accuracy and computational cost. Explicitly, the leapfrog integrator is given by

$$\begin{aligned}
\boldsymbol{u}(t + \Delta t/2) &= \boldsymbol{u}(t - \Delta t/2) + \boldsymbol{\nabla}\Phi(t)\,\Delta t\ , \tag{4.16}\\
\boldsymbol{r}(t + \Delta t) &= \boldsymbol{r}(t) + \boldsymbol{u}(t + \Delta t/2)\,\Delta t\ . \tag{4.17}
\end{aligned}$$

4.2 Improvements: P^{3}M, TreePM and AMR

Although being a significant improvement over the direct summation approach in terms of computational efficiency, the PM algorithm suffers from several drawbacks. The small-scale resolution of a PM simulation is limited by the size of the cells of the Fourier mesh; structure on scales smaller than the grid spacing is smoothed out. Since the memory consumption scales as N_{g}^3, it quickly becomes infeasible to simulate large boxes with sufficient small scale resolution. Furthermore, the gravitational force is anisotropic on scales comparable to the grid constant, which can cause artifacts in structures slightly larger than the grid spacing.

These problems are solved in the P^{3}M algorithm (Hockney & Eastwood 1988) by splitting the gravitational potential of a particle into a short-range and a long-range component:

$$\Phi(r) = \Phi^{\mathrm{l}}(r) + \Phi^{\mathrm{s}}(r)\ , \tag{4.18}$$

where

$$\Phi^{\mathrm{l}}(r) \;=\; -\frac{Gm_{\mathrm{p}}}{r}\,\mathrm{erf}\left(\frac{r}{2r_{\mathrm{s}}}\right) \tag{4.19}$$

$$\Phi^{\mathrm{s}}(r) \;=\; -\frac{Gm_{\mathrm{p}}}{r}\,\mathrm{erfc}\left(\frac{r}{2r_{\mathrm{s}}}\right)\;. \tag{4.20}$$

Here, m_{p} is the particle mass, erf and erfc are the error function and its complement, and r_{s} is the split radius. Since the long range potential is very smooth, it can still be computed on the PM grid. The short range potential is only important for particle separations smaller than $\approx 4r_{\mathrm{s}}$, meaning that all particles more distant than this from a given point will not contribute to Φ_{s} at that point. Therefore, one constructs a sphere around each particle within which the short range contribution to the potential at the particle position is computed by direct summation over all other particles in the sphere. Therefore, the small scale resolution is no longer limited by the FFT mesh. At the same time, the computational burden is significantly decreased compared to the brute force approach by the possibility to compute the contribution to Φ from remote particles using the FFT.

In the P³M algorithm, a new issue arises due to the finite sampling of the phase space with particles of large masses. If one used the gravitational potential of a point mass to compute the short range potential, hard binary systems would form and scattering events with very large deflections of the particle trajectories would occur. This is in contradiction with the collisionless nature of dark matter. To avoid this unphysical behaviour, the gravitational potential of a particle is *softened* (Hockney & Eastwood 1988; Dehnen 2001) according to

$$\Phi_{\mathrm{soft}}(r) = \frac{Gm_{\mathrm{p}}}{r+\epsilon}\;, \tag{4.21}$$

where ϵ is the *softening length*. Since the potential approaches a constant value for $r \lesssim \epsilon$, the force between two particles tends to zero. Therefore, even in high-density regions the resolution of the simulation is limited to scales above $\approx 2\epsilon$. Note that this problem is not present in the PM algorithm because of its inherent smoothing due to the FFT mesh.

However, the P³M algorithm is not in widespread use for cosmological simulations. The reason for this is that when non-linear clustering starts in earnest, the number of particles contributing to Φ_{s} at a given point can still become very large, and so the simulation stalls due to the inefficiency of the direct potential computation. A remedy to this problem is provided by the TreePM algorithm, where the direct summation is replaced with a tree-based procedure. A tree code combines particles into a hierarchy of cells. The computation is sped up by grouping more distant particles together and treating them as a new, more massive pseudo-particle. This reduces the number of summations significantly, at the cost of introducing a small error due to the averaging involved. More details can be found in, e.g., Springel (2005).

A different approach to improve the small-scale resolution of the PM algorithm is *adaptive mesh refinement* (AMR), as implemented for example in `ART` (Gottlöber & Klypin 2008, and references therein). Here, the computation of the potential is always carried out on a grid. However, grid cells in high-density regions can be recursively subdivided into smaller cells, hence increasing spatial resolution where needed. This leads to a hierarchy of grid refinements, the geometry of which fully adapts to the density field. The FFT can only be used to compute the potential on the zeroth-order refinement, which is always a regular cubic grid. This provides boundary conditions for the irregular regions where the grid has been refined. The potential for the first refinement level is then obtained by solving the Dirichlet boundary value problem. This solution, in turn, provides boundary conditions for the next higher refinement level, and so on.

4.3 Halo identification

A generic feature of cosmological structure formation is the formation of well-defined, self-bound structures, so-called dark matter halos. Since these halos and possible substructures within the halos mark the location of galaxies, galaxy groups and clusters, their identification in a N-body simulation is of great interest.

Several algorithms exist to identify dark matter halos. The probably most common approach is the *Friends-of-Friends* (FoF) algorithm (Davis et al. 1985). After specifying a linking length l (typically 20% of the mean separation of two particles in the simulation), particles are said to belong to a particular halo if they are not further away from a halo member than the linking length.

The halos formed by the FoF algorithm approximately correspond to regions within the isodensity contour with density $\propto 1/l^3$ (Springel et al. 2001). The standard FoF technique has several shortcomings: it cannot detect substructures within halos and tends to artificially link together physically distinct objects. The hierarchical FoF algorithm by Gottlöber et al. (1999) addresses these problems by reducing the linking length in several steps, subsequently revealing substructures of higher density contrast. However, there still is a considerable false detection rate, in particular in high-density regions. Therefore, in the DENMAX (Bertschinger & Gelb 1991) and SKID (Governato et al. 1997) algorithms, the geometrical information from the FoF algorithm is complemented with dynamical information to help decide whether a halo is real and to remove unbound particles from the halo. Additional information can be obtained by checking whether the candidate halo also exists at later times in the simulation. The SUBFIND algorithm (Springel et al. 2001), which has been used to construct the halo catalogues of the Millennium Simulation, starts with the groups detected by the FoF method. A local density estimate is computed at the position of each particle in the FoF halo, and local overdensities are found by looking for isodensity contours that pass through a saddle point of the estimated density. Finally, also in this algorithm unbound particles are removed based on dynamical information.

Figure 4.1: *An example for a merger history (merger tree) of a dark matter halo (taken from Stewart et al. 2008). The resulting halo has a mass of $3.1 \times 10^{12}\,M_\odot$ at z=0. Solid black lines correspond to isolated field halos, dashed red lines to subhalos that merge with the main halo.*

The algorithms described do not only result in halo and subhalo catalogs for each snapshot redshift, but also allow the reconstruction of the merger histories of all halos identified in the simulation at $z = 0$. These *merger trees* (an example is given in Fig. 4.1) are a key ingredient to the semi-analytic modelling of galaxy formation, as is described in the next section.

4.4 Semi-analytic models of galaxy formation

It is not yet possible to satisfactorily simulate the formation of galaxies in a cosmological volume. On the one hand this is because of the computational resources necessary to fully treat the baryonic physics and the dark matter simultaneously, and on the other hand because the physics of galaxy formation is not yet fully understood. Semi-analytic models of galaxy formation (SAMs) deal with these difficulties by combining a simple analytical parametrization of galaxy formation with the merger trees obtained numerically for halos and subhalos in *N*-body simulations. This approach has the advantage that the effect of variations in the model parameters can be investigated with comparably little computational effort, because it only relies on the dark matter merger trees. By matching the model to observations it is then possible to identify physical processes that lead to a certain property of the galaxy distribution and to assess their relative importance.

Since we will use the resulting mock galaxy catalogues in Chapter 7 for our simulations of galaxy-galaxy-lensing, we now give a simplified description of the model that has been implemented by Croton et al. (2006) and De Lucia & Blaizot (2007) for the Millennium Simulation:

- When a dark matter halo forms, a certain mass fraction of baryons will collapse with it. This fraction is assumed to be the global baryon fraction in the Universe. The gas initially has primordial composition, but can at later times be enriched by stellar mass loss.

- The gas in the halo cools by radiation. The cooling rate, i.e. the rate of change of the cold gas mass in the halo, is determined by the metallicity of the gas and the mass of the dark matter halo. Cool gas is assumed to form a disk at the center of the halo.

- Star formation takes place in the cool gas disk, either quiescently or in bursts:
  - For quiescent star formation to occur, it is necessary for the gas disk to exceed a certain surface mass density, or, assuming a homogeneous cold disk, a critical cold gas mass. This results in an estimate for the star formation rate. The efficiency of star formation in the disk can be controlled by a model parameter.
  - Starbursts occur when two galaxy host halos merge. A certain, adjustable fraction of the total cold gas mass of the merging halos is converted into stars.

- Several feedback processes that reheat the gas in a halo can be accounted for:
  - A certain fraction of newly formed stars will explode as supernovae. These inject energy into the halo and heat the gas. The rate of change of the hot gas mass is assumed to be proportional to the mass of stars formed over some finite time interval (and thus to the supernova rate). In low-mass halos, it is possible that some gas is ejected from the halo, either to fall back later or to be lost permanently.
  - Supermassive black holes at the centers of halos accrete gas and release some of the energy of the infalling matter as radiation. This can occur in the "radio mode", where steady, quiescent accretion takes place, or in the more violent "quasar mode", which is activated in halo mergers. In either case this counteracts the gas cooling and converts cool gas to hot gas.

- The morphology of galaxies (i.e. the sizes of disks and bulges) is determined by disc growth through accretion, bulge growth due to disk instabilities and bulge formation through mergers. Mergers of galaxies occur in the model between halo or subhalo central galaxies

Parameter	Millennium Simulation	WMAP-5
$L \, [h^{-1}\,\mathrm{Mpc}]$	500	150
N	2160^3	256^3
$m_\mathrm{p} \, [h^{-1}\,M_\odot]$	8.6×10^8	1.2×10^{10}
$\epsilon \, [h^{-1}\,\mathrm{kpc}]$	5.0	20.0
Ω_m	0.25	0.25
Ω_Λ	0.75	0.75
Ω_b	0.045	0.04
σ_8	0.9	0.78
h_{100}	0.73	0.7
n_s	1.0	1.0
Transfer function	`CMBfast`[1]	Eisenstein & Hu (1999)

Table 4.1: *Parameters of the simulations used in this thesis. The side length of the simulation volume is denoted by L, the number of particles by N, the particle mass by m_p and the smoothing length by ϵ. All lengths given are comoving.*

and satellite galaxies in the same (sub)halo which have lost their own subhalo. This happens when a subhalo is tidally stripped and eventually is below the detection threshold of the halo finding algorithm. The orphan galaxies merge with the central galaxy after a time set by the dynamical friction timescale. If the merger is an unequal-mass merger, then the stars of the satellite galaxies are transferred to the bulge of the central galaxy, while the cold gas is added to the disk. If the masses of the two galaxies are comparable, then both discs are destroyed and all stars end up in the spheroidal component.

- The spectral properties of the galaxies can be modelled using population synthesis, since the ages of the different stellar populations and their metallicities are known.

This recipe leads to a system of coupled differential equations for the galaxy properties (such as gas and stellar masses) which can be solved numerically, traversing the merger tree from early to late times. The result is a mock galaxy catalogue that contains besides observable properties of the galaxies such as magnitudes in various filter bands also the properties of their host halos.

Finally, it should be noted that the galaxies in a SAM are massless point particles (usually the central particles of dark matter halos and subhalos), whose presence does not alter the matter distribution in the N-body simulation in any way. Therefore, within the framework of these models it is not possible to study, e.g., the effect of baryonic physics on the matter power spectrum.

4.5 N-body simulations used in this thesis

Finally, we briefly describe the simulations used in this thesis. The main parameters are also summarized in Tab. 4.1.

4.5.1 The Millennium Simulation

The Millennium Simulation (Springel et al. 2005) is one of the largest simulations of structure formation so far. It was run using a modified version of the publically available TreePM code

---

[1] Seljak & Zaldarriaga (1996)

`GADGET-2` (Springel 2005). It tracks the evolution of 2160^3 ($\approx 10^{10}$) dark matter particles from $z = 127$ to $z = 0$. Snapshots of the particle distribution are provided at 64 redshifts within this interval. The size of the simulation box is $L = 500 \, h^{-1}$ Mpc, which is big enough to contain rare massive objects, while the large number of particles and their low mass of $8.6 \times 10^8 \, h^{-1} \, M_\odot$ allow one to resolve structures equivalent the size of the Small Magellanic Cloud. These structures and their substructures were identified using the `SUBFIND` algorithm. The price paid for this large dynamic range is the data volume: storage of all 64 snaphots requires ≈ 20 TB of disk space. A large amount of the simulation data, such as the halo and subhalo catalogs, and also the results of the galaxy formation model by De Lucia & Blaizot (2007), have been made publically available through an online database (Lemson & The Virgo Consortium 2006), which we have made extensive use of for Chapters 5, 7 and 8.

4.5.2 Smaller simulations

We have run a set of 10 smaller N-body simulations (CDFS simulations henceforth) using the public version of `GADGET-2`. All simulations are realizations of the same WMAP-5-like cosmology ($\Omega_{\rm m} = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_{\rm b} = 0.04$, $n_{\rm s} = 1.0$, $\sigma_8 = 0.78$, $h_{100} = 0.7$). The simulation boxes are $L_{\rm box} = 150 \, h^{-1}$Mpc on a side, populated by $N_{\rm p} = 256^3$ dark matter particles with masses of $m_{\rm p} = 1.2 \times 10^{10} \, h^{-1} \, M_\odot$. We have started the simulations at $z = 50$ and obtained snapshots from $z = 0$ to $z = 4.5$ in intervals of Δz corresponding to the box size. The initial particle distribution was created by perturbing a glass distribution using the Zel'Dovich approximation. Halos were identified using a FoF code kindly provided by V. Springel.

5 Ray-Tracing Simulations

Analytical predictions for weak gravitational lensing necessarily involve simplifications, such as the Born approximation where the deflections of the light rays are neglected. Depending on the context, these approximations may not be accurate enough; as we have seen in Sec. 3.3, in some situations correction terms for these approximations can be derived. However, all analytical predictions currently are based in one way or another on fitting formulae to N-body simulations, and therefore have a limited accuracy (currently $\approx 10\%$ for two-point statistics; the uncertainties can be considerably larger for higher-order measures). Thus, whenever higher precision is needed or the problem at hand simply evades an analytical description, one has to resort to ray-tracing simulations, where the propagation of light rays through the (suitably discretized) large-scale structure is studied explicitly. These simulations can also be coupled with semi-analytic models of galaxy formation, so that it is possible to obtain realistic mock observations including effects of the galaxy sample selection. Recent efforts in this field were undertaken by e.g. Jain et al. (2000) or Vale & White (2003).

The actual implementation of a ray-tracing simulation depends on the properties of the N-body simulation from which the matter distribution along the line of sight (l.o.s) is constructed. In this section, we first describe what we will call the "standard" ray-tracing algorithm, which is suited for work with N-body simulations with smaller boxes ($L \lesssim 200 \, h^{-1}$ Mpc). Our implementation of this method, called `RTsuite`, has already been discussed in detail in Hartlap (2005), and simulations produced with `RTsuite` have been used in Dietrich et al. (2007), Eifler et al. (2008), Pielorz (2008) and Chapter 6 of this thesis. We then describe a code that we have developed for the Millennium Simulation, whose large simulation volume and at the same time excellent small scale resolution require some modifications to the standard ray-tracing algorithm. The code has been thoroughly tested and applied to some aspects of weak gravitational lensing in Hilbert et al. (2008a). Furthermore, it has been used to study the statistics of strong lensing events (Hilbert et al. 2007b, 2008b) and for an investigation of possible benefits of using the 21 cm radiation of pre-galactic hydrogen in Hilbert et al. (2007a).

5.1 The multiple-lens-plane algorithm

We first consider the standard implementation of the multiple-lens-plane algorithm (see, e.g., Blandford & Narayan 1986; Schneider et al. 1992; Seitz et al. 1994; Jain et al. 2000), which basically comprises a discretization of Eq. (3.60).

The first step towards simulating the propagation of light rays in an inhomogeneous universe is to construct the density field along the backwards lightcone of the observer. The matter distribution is taken from N-body simulations, which provide *snapshots* of the simulation volume at different redshifts. If the side length of the simulation volume is not too large ($L \lesssim 200 \, h^{-1}$ Mpc), the density along the l.o.s. from the observer can be constructed in a straight-forward way: the snapshots are lined up along the line of sight (which is assumed to be parallel to one of the coordinate axes of the snapshots), using snaphots of higher redshift as the distance to the observer increases. However, measures have to be taken to avoid the repetition of the same structures

along the l.o.s.: a snapshot for $z = z_1$ contains essentially the same matter distribution as a snapshot for $z_2 > z_1$, even though at a slightly different stage of evolution. If the snapshots were simply lined up, the light rays would encounter the same objects every time after covering a distance of $\approx L$. Therefore, one makes use of the periodic boundary conditions of the N-body simulation and randomly applies rotations by 90° about the coordinate axes, translations and and parity transformations to the particle distribution in the snapshots.

The matter distribution of each snapshot is then projected plane-parallely along one of the coordinate axes onto a *lens plane*, and light deflections are assumed only to occur when a light ray intersects one of these planes. The matter distribution on the lens planes is periodic, which is a consequence of the periodic boundary conditions of the N-body simulation. This allows for the computation of the deflection angle and its derivatives from the projected density on each plane by two-dimensional FFT methods.

After these preparations, the light rays are shot into the stack of lens planes, starting at the observer. This is usually done in a way such that the light rays form a uniform grid on the first lens plane (the "sky" seen by the observer). The light rays are propagated from lens plane to lens plane using the precomputed deflection angles on the lens planes (interpolated onto the ray positions) and a recursion formula for the angular position of each ray. A similar recursion relation also yields the Jacobian matrix of the lens mapping from the observer to the lens planes. With this, the effective convergence κ and the shear γ can be computed at the positions of the light rays. Several (quasi-)independent realizations of the convergence and shear fields can be obtained by choosing a different observer position with respect to the first lens plane and/or by repeating the simulation with different random transformations applied to the snapshots before projection onto the lens planes.

Besides the discretization of the matter distribution, the multiple-lens-plane algorithm involves several approximations:

- It is assumed that the density field does not evolve much during the time a light ray needs to traverse a snapshot.

- The geometrical weight function G in Eq. (3.60) is assumed to be a slowly changing function of redshift, and therefore can be assumed to be constant within a snapshot.

- The projection is done plane-parallely. The validity of this approximation implies that the angles enclosed by the light rays with the l.o.s. are small, limiting the angular size of a simulated patch of the sky to $\approx 10° \times 10°$. This is also called the *flat sky approximation*.

- Eq. (3.60) requires the gradient of the gravitational potential to be computed perpendicular to the light ray. In the multiple-lens-plane algorithm, this gradient is computed in the plane of the sky, however. This is valid within the flat sky approximation, with the additional assumption that the deflections suffered by the light rays are small.

We will now discuss the particular steps of the multiple-lens-plane algorithm in more detail. To simplify the notation, we introduce the following abbreviations: for the comoving angular diameter distance $f_K(w^{(k)})$ to the k-th lens plane we write $f_K^{(k)}$, and the distance $f_K(w^{(k)} - w^{(l)})$ between plane k and plane l ($k > l$) is abbreviated by $f_K^{(l,k)}$. Furthermore, $a^{(k)} = a(w^{(k)})$.

5.1.1 Construction of the lens planes

To construct the lens planes, the particles in a particular snapshot are projected along one of the coordinate axes onto a two-dimensional grid, using one of the assignment schemes described

in Sec. 4.1.2. We parametrize the projected matter distribution on the k-th lens plane by the *dimensionless surface mass density* $\sigma^{(k)}$:

$$\sigma^{(k)}(\boldsymbol{\beta}^{(k)}) = \frac{3H_0^2 \Omega_{\mathrm{m}}}{2c^2} \frac{f_K^{(k)}}{a^{(k)}} \int_{w_{\mathrm{L}}^{(k)}}^{w_{\mathrm{U}}^{(k)}} \mathrm{d}w' \, \delta(\boldsymbol{\beta}^{(k)}, w') \;. \tag{5.1}$$

Here, $\boldsymbol{\beta}^{(k)}$ is the angular position on the lens plane and δ the three-dimensional matter density contrast as defined in Sec. 2.2.2. The integration limits $w_{\mathrm{L}}^{(k)}$ and $w_{\mathrm{U}}^{(k)}$ define the lower and upper boundaries, respectively, of the slice of the matter distribution which is projected onto the k-th lens plane (with $w_{\mathrm{U}}^{(k)} = w_{\mathrm{L}}^{(k+1)}$). In the standard implementation of the ray-tracing algorithm, $w_{\mathrm{L}}^{(k)}$ and $w_{\mathrm{U}}^{(k)}$ are simply given by the boundaries of the simulation volumes; as we will discuss further below, we have to make a different choice for the Millennium Simulation. The positions of the lens planes within the slices are usually chosen to correspond to the mean redshifts (e.g. Jain et al. 2000) or mean comoving distances (e.g. Wambsganss et al. 2004) of the slice boundaries. The exact choice becomes unimportant for sufficiently small spacings between the lens planes.

The deflection $\boldsymbol{\alpha}^{(k)}(\boldsymbol{\beta}^{(k)})$ of a light ray intersecting the k-th lens plane at angular position $\boldsymbol{\beta}^{(k)}$ can be expressed as the gradient of a *lensing potential* $\psi^{(k)}$ (see Eq. 3.8):

$$\boldsymbol{\alpha}^{(k)}(\boldsymbol{\beta}^{(k)}) = \boldsymbol{\nabla}_{\boldsymbol{\beta}^{(k)}} \psi^{(k)}(\boldsymbol{\beta}^{(k)}) \;, \tag{5.2}$$

which is related to the dimensionless surface mass density through the Poisson equation:

$$\boldsymbol{\nabla}^2_{\boldsymbol{\beta}^{(k)}} \psi^{(k)}(\boldsymbol{\beta}^{(k)}) = 2\sigma^{(k)}(\boldsymbol{\beta}^{(k)}) \;. \tag{5.3}$$

For later use, we also define the *shear matrix* for the k-th lens plane

$$\mathsf{U}_{ij}^{(k)} = \frac{\partial^2 \psi^{(k)}(\boldsymbol{\beta}^{(k)})}{\partial \beta_i^{(k)} \partial \beta_j^{(k)}} = \frac{\partial \alpha_i^{(k)}(\boldsymbol{\beta}^{(k)})}{\partial \beta_j^{(k)}} \;. \tag{5.4}$$

Since the matter distribution on a lens plane is periodic, it is convenient to compute the deflection potential, the deflection angle and the shear matrix in Fourier space:

$$\tilde{\psi}^{(k)}(\boldsymbol{\ell}) \;=\; -2\frac{\tilde{\sigma}^{(k)}(\boldsymbol{\ell})}{|\boldsymbol{\ell}|^2} \;, \tag{5.5}$$

$$\tilde{\boldsymbol{\alpha}}^{(k)}(\boldsymbol{\ell}) \;=\; -\mathrm{i}\boldsymbol{\ell}\,\tilde{\psi}^{(k)}(\boldsymbol{\ell}) = 2\mathrm{i}\,\tilde{\sigma}^{(k)}(\boldsymbol{\ell})\,\frac{\boldsymbol{\ell}}{|\boldsymbol{\ell}|^2} \;, \tag{5.6}$$

$$\tilde{\mathsf{U}}_{ij}^{(k)} \;=\; -\ell_i\ell_j\,\tilde{\psi}^{(k)}(\boldsymbol{\ell}) = \frac{2\ell_i\ell_j}{|\boldsymbol{\ell}|^2}\,\tilde{\sigma}^{(k)}(\boldsymbol{\ell}) \;, \tag{5.7}$$

for $\boldsymbol{\ell} \neq 0$. This allows for a fast computation of the deflection angles and the shear matrices from the surface mass density using the Fast Fourier Transform.

5.1.2 Ray-Tracing

Given the deflection angles on the lens planes, we can compute the angular position $\boldsymbol{\beta}^{(k)}$ of a particular light ray (reaching the observer from angular position $\boldsymbol{\theta} \equiv \boldsymbol{\beta}^{(1)}$) using the generalized lens equation

$$\boldsymbol{\beta}^{(k)}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \sum_{i=1}^{k-1} \frac{f_K^{(i,k)}}{f_K^{(k)}} \boldsymbol{\alpha}^{(i)}(\boldsymbol{\beta}^{(i)}) \;, \quad k = 2, 3, \ldots \tag{5.8}$$
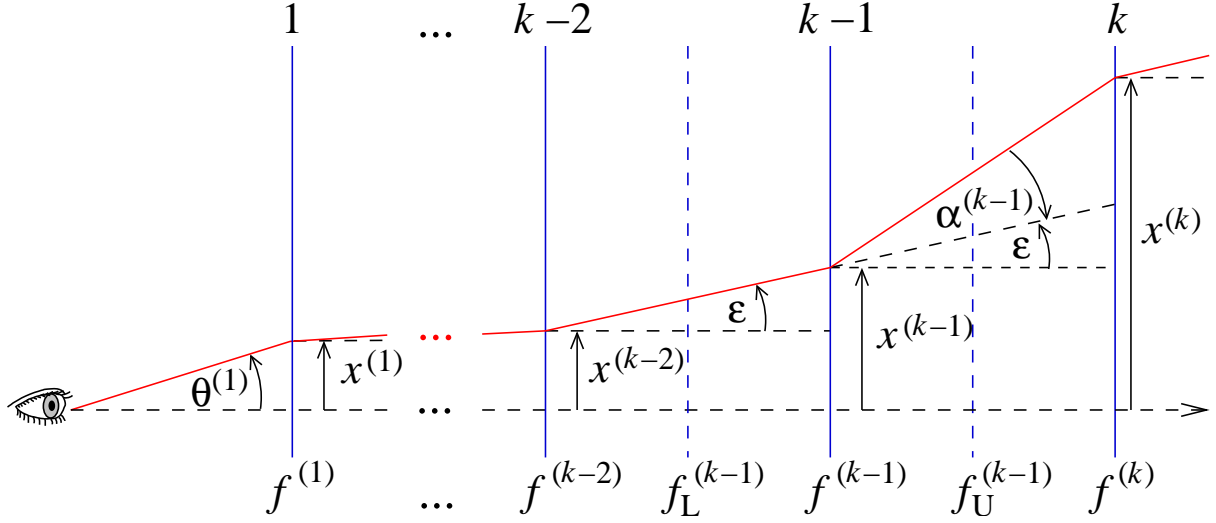
Figure 5.1: *Schematic view of the observer's backward light cone in the multiple-lens-plane approximation. A light ray (red line) experiences a deflection only when passing through a lens plane (solid blue lines). The deflection angle $\boldsymbol{\alpha}^{(k-1)}$ of a ray passing through the lens plane at distance $f_{\mathrm{L}}^{(k-1)}$ from the observer is obtained from the matter distribution between $f_{\mathrm{U}}^{(k-1)}$ and $f_{\mathrm{L}}^{(k-1)}$ projected onto the plane. Using the deflection angle $\boldsymbol{\alpha}^{(k-1)}$ of the light ray at the previous lens plane and the ray's positions $\boldsymbol{x}^{(k-1)}$ and $\boldsymbol{x}^{(k-2)}$ on the two previous planes, the position $\boldsymbol{x}^{(k)}$ on the current plane can be computed (figure by S. Hilbert).*

In the case of $k = 2$, this reduces to the ordinary lens equation (Eq. 3.4). Eq. (5.8) defines a series of lens mappings from the image plane (i.e. the first lens plane) to the other lens planes:

$$\boldsymbol{\theta} \mapsto \boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^{(k)}(\boldsymbol{\theta}) \, , \quad k = 1, 2, \dots \tag{5.9}$$

Differentiating Eq. (5.8) with respect to $\boldsymbol{\theta}$, we obtain a corresponding equation for the Jacobian of the lens mapping to the k-th lens plane :

$$\mathsf{A}^{(k)}(\boldsymbol{\theta}) = \mathsf{E} - \sum_{i=1}^{k-1} \frac{f_K^{(i,k)}}{f_K^{(k)}} \, \mathsf{U}^{(i)}(\boldsymbol{\beta}^{(i)}) \, \mathsf{A}^{(i)}(\boldsymbol{\theta}) \, . \tag{5.10}$$

Since Eqs. (5.8) and (5.10) require the storage of the ray positions and Jacobians on all previous lens planes, it is not practical for tracing rays through many lens planes at high resolution.

An alternative expression can be obtained by the following geometrical consideration (see, e.g., Schneider et al. 1992; Seitz et al. 1994; Hartlap 2005): The transverse comoving position $\boldsymbol{x}^{(k)}$ of a light ray on the lens plane k is related to the positions $\boldsymbol{x}^{(k-2)}$ and $\boldsymbol{x}^{(k-1)}$ on the two previous lens planes by (see Fig. 5.1):

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + f_K^{(k-1,k)} \left[\boldsymbol{\epsilon} - \boldsymbol{\alpha}^{(k-1)}(\boldsymbol{x}^{(k-1)}) \right], \text{ where}$$
$$\boldsymbol{\epsilon} = \frac{\boldsymbol{x}^{(k-1)} - \boldsymbol{x}^{(k-2)}}{f_K^{(k-2,k-1)}} \, . \tag{5.11}$$

Using $\boldsymbol{x}^{(k)} = f_K^{(k)} \boldsymbol{\beta}^{(k)}$, this can be written in terms of angular coordinates:

$$
\boldsymbol{\beta}^{(k)} = \frac{1}{f_K^{(k)}} \left[\left(1 + \frac{f_K^{(k-1,k)}}{f_K^{(k-2,k-1)}} \right) f_K^{(k-1)} \boldsymbol{\beta}^{(k-1)} \right.
$$
$$
\left. - f_K^{(k-1,k)} \boldsymbol{\alpha}^{(k-1)}(\boldsymbol{\beta}^{(k-1)}) - \frac{f_K^{(k-1,k)}}{f_K^{(k-2,k-1)}} f_K^{(k-2)} \boldsymbol{\beta}^{(k-2)} \right] . \tag{5.12}
$$

For a light ray reaching the observer from angular position $\boldsymbol{\theta}$ on the first lens plane, we can compute its angular position on the other lens planes by iterating Eq. (5.12) with the initial values $\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}^{(1)} = \boldsymbol{\theta}$. In App. A, we give an alternative derivation of this equation by directly discretizing the differential equation for the comoving separation vector between two light rays (Eq. 3.57). Differentiating Eq. (5.12) with respect to $\boldsymbol{\theta}$, we obtain a recurrence relation for the Jacobian matrix of the lens mapping (Eq. 5.9):

$$
\mathsf{A}^{(k)}(\boldsymbol{\theta}) = \left(1 + \frac{f_K^{(k-1,k)}}{f_K^{(k-2,k-1)}} \right) \frac{f_K^{(k-1)}}{f_K^{(k)}} \mathsf{A}^{(k-1)}(\boldsymbol{\theta})
$$
$$
- \frac{f_K^{(k-1,k)}}{f_K^{(k)}} \mathsf{U}^{(k-1)}(\boldsymbol{\beta}^{(k-1)}) \mathsf{A}^{(k-1)}(\boldsymbol{\theta}) - \frac{f_K^{(k-1,k)}}{f_K^{(k-2,k-1)}} \frac{f_K^{(k-2)}}{f_K^{(k)}} \mathsf{A}^{(k-2)}(\boldsymbol{\theta}) . \tag{5.13}
$$

With the initial values $\mathsf{A}^{(0)} = \mathsf{A}^{(1)} = \mathsf{E}$, this equation allows us to iteratively compute the Jacobian for a particular light ray from the observer to any lens plane. It requires in practice much fewer arithmetic operations and memory than the commonly used recurrence relations (e.g. by Jain et al. 2000) based on Eq. (5.8).

5.1.3 Construction of mock observations

From Eq. (5.13), we can compute the Jacobian matrix for any given source redshift: we denote with $f_K^{(k+\lambda)}$ the comoving angular diameter distance to the source plane, where $\lambda \in [0,1]$ and $f_K^{(k)} \leq f_K^{(k+\lambda)} \leq f_K^{(k+1)}$, and with $\mathsf{A}^{(k+\lambda)}(\boldsymbol{\theta})$ the corresponding Jacobian matrix. The latter can be computed from $\mathsf{A}^{(k)}$ and $\mathsf{A}^{(k-1)}$ using the recursion relation Eq. (5.13):

$$
\mathsf{A}^{(k+\lambda)}(\boldsymbol{\theta}) = \left(1 + \frac{f_K^{(k,k+\lambda)}}{f_K^{(k-1,k)}} \right) \frac{f_K^{(k)}}{f_K^{(k+\lambda)}} \mathsf{A}^{(k)}(\boldsymbol{\theta})
$$
$$
- \frac{f_K^{(k,k+\lambda)}}{f_K^{(k+\lambda)}} \mathsf{U}^{(k)}(\boldsymbol{\beta}^{(k)}) \mathsf{A}^{(k)}(\boldsymbol{\theta}) - \frac{f_K^{(k,k+\lambda)}}{f_K^{(k-1,k)}} \frac{f_K^{(k-1)}}{f_K^{(k+\lambda)}} \mathsf{A}^{(k-1)}(\boldsymbol{\theta}) . \tag{5.14}
$$

Subtracting this expression from the analogous one for $\mathsf{A}^{(k+1)}$, the terms involving the shear matrices and $\mathsf{A}^{(k-1)}$ cancel out. Rearranging the remaining terms, we find

$$
\mathsf{A}^{(k+\lambda)}(\boldsymbol{\theta}) = \frac{f_K^{(k,k+\lambda)}}{f_K^{(k+\lambda)}} \frac{f_K^{(k+1)}}{f_K^{(k,k+1)}} \mathsf{A}^{(k+1)}(\boldsymbol{\theta}) + \frac{f_K^{(k)}}{f_K^{(k+\lambda)}} \left(1 - \frac{f_K^{(k,k+\lambda)}}{f_K^{(k,k+1)}} \right) \mathsf{A}^{(k)}(\boldsymbol{\theta}) . \tag{5.15}
$$

With this, we can obtain the shear, convergence, etc. on a grid using Eq. (3.33). We prefer the definition of κ and γ by the polar decomposition of A over the definition using the decomposition according to spin (Eq. 3.38). This is because κ' and γ' as defined by Eq. (3.34) can in general
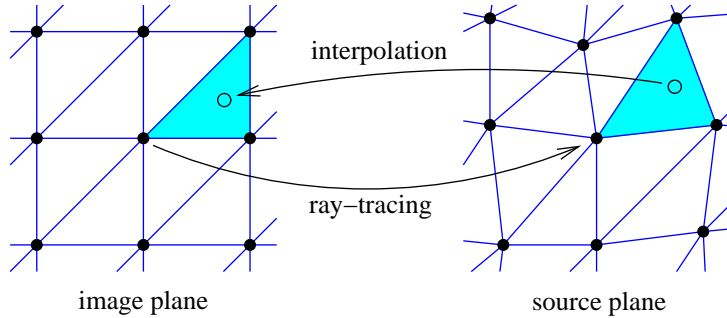
Figure 5.2: *Interpolation scheme used for determining image positions of galaxies. The regular grid of rays in the image plane (left filled circles) is used to partition the image plane into triangles (right blue lines). The image position (left open circle) of a source inside a triangle (left blue lines) formed by the backtraced rays on the source plane (right filled circles) is then determined by linear interpolation (figure by S. Hilbert).*

not be related to observables in a simple way, whereas the unprimed quantities in Eq. (3.33) are directly related to the observed image distortions.

If the source galaxies are distributed in redshift according to the redshift distribution function $p_z(z)$ (or, equivalently with a distribution in comoving distance $p_w(w) = p_z(z) |\mathrm{d}z/\mathrm{d}w|$), we have to compute shear and convergence from an average of the Jacobian matrices for different redshifts. We note that, given the discretization scheme of the multiple-lens-plane algorithm, the most sensible location for the source galaxies are the back sides of the slices of the matter distribution that are projected onto the lens planes (i.e., at $w_{\mathrm{U}}^{(k)}$). The reason for this is that the deflections at a lens plane account for the effects of all matter in the respective slice. If a source galaxy were located somewhere within the slice, the shear assigned to it with the procedure outlined above would therefore contain contributions from matter that is actually at higher z than the source. To avoid this kind of conceptual difficulties, we use the Jacobians $\mathsf{A}^{(k+1/2)}$ for the lens mapping to the back sides of the slices (obtained from Eq. 5.15) to compute the average over the source galaxy redshift distribution:

$$\mathsf{A}(\boldsymbol{\theta}) = \sum_{k=1}^{k_{\max}} p_w(w^{(k)}) \, \mathsf{A}^{(k+1/2)}(\boldsymbol{\theta}) \left(w^{(k+1)} - w^{(k)} \right) \ . \tag{5.16}$$

Here, $k_{\max}$ is given by the redshift of the most distant source. A generic distribution, which we will frequently make use of in this thesis, was proposed by Smail et al. (1995) and is given by

$$p(z) = A \left(\frac{z}{z_0} \right)^{\alpha} \exp \left[- \left(\frac{z}{z_0} \right)^{\beta} \right] \ , \tag{5.17}$$

where A is a normalization constant and α, β and z_0 are free parameters.

5.1.4 Including galaxies from semi-analytic models

In order to realistically model galaxy-galaxy lensing, we need a method to include galaxies from semi-analytical models of galaxy formation in our simulation. In particular, the task is to translate the galaxy positions on the lens planes into observed, lensed positions on the sky. We therefore have to compute the image positions $\boldsymbol{\theta}_{\mathrm{g}}$ given the galaxies' source positions $\boldsymbol{\beta}_{\mathrm{g}}^{(k)}$ (i.e. the positions on the lens planes). Note that this is the exact opposite of the problem we have solved with the multiple-lens-plane algorithm, where we have computed the ray positions on the lens planes given the ray positions on the image plane.

During the ray-tracing, we store the positions of the light rays on every lens plane. We partition the region of the image plane that is covered by the grid of rays into triangles formed by rays of adjacent grid points (Schneider et al. 1992), as is illustrated in Fig. 5.2. Using the stored ray positions, we are able to "backtrace" these triangles to the lens planes. On each lens plane, we then identify for each such triangle all galaxies which are located inside the backtraced triangle. The image positions of these galaxies and the magnifications at their positions are then computed by linear interpolation of the Jacobian matrices of the three rays forming the triangle. If a particular galaxy is covered by more than one triangle, the galaxy is multiply imaged; the method therefore can to a certain degree also account for strong lensing (see the discussion in Schneider et al. 1992).

Furthermore, gravitational lensing affects the observed magnitudes of the galaxies through the magnification effect. We take this into account by correcting the apparent magnitude in the filter band X according to

$$m_X^{(\text{lensed})} = m_X^{(\text{unlensed})} - 2.5 \log(\mu) \,, \tag{5.18}$$

where μ is the previously computed magnification.

5.1.5 Ray-Tracing with the Millennium Simulation

We now discuss several modifications of the standard ray-tracing algorithm outlined in the previous section which are necessary to deal with high-resolution N-body simulations with large simulation boxes, such as the Millennium Simulation. These affect in particular the construction of the lens planes; the actual ray-tracing part of the algorithm can be used essentially as described in Sec. 5.1.2.

Construction of the lens planes

With a comoving size of $L = 500 \, h^{-1}$ Mpc, the simulation box of the Millennium Simulation is too big to be projected onto a single lens plane; doing so would result in an unacceptably large discretization error. Neither can the evolution of the matter distribution within the light travel time through the box be neglected, nor can the geometrical weight factors in Eq. (3.60) be approximated to be constant. Therefore, it is necessary to construct the lens planes from slices that are smaller than the simulation volume. This immediately raises the issue of how to deal with dark matter halos that are located close to or even on a slice boundary. Simply slicing through halos would result in one part of a galaxy cluster to appear on the k-th lens plane, and the remaining part on the $k + 1$-st plane, spreading the halo over a length of the order of $\sim 100 \, h^{-1}$ Mpc. For this reason, the slice boundaries have to be adaptive so that halos are either completely included or excluded from a slice. Furthermore, large-scale correlations that extend beyond the slice boundaries have to be taken into account and to be preserved in the slicing procedure. Neglecting any of these issues would detract from the goal of making the best possible use of the large dynamic range of the Millennium Simulation. Therefore, the method for the construction of the lens planes described in the previous section cannot be used here. The random transformations of the matter distribution in the slices, applied to avoid the repetition of structures along the l.o.s., would be incompatible with the adaptive slice boundaries and destroy any large-scale correlations.
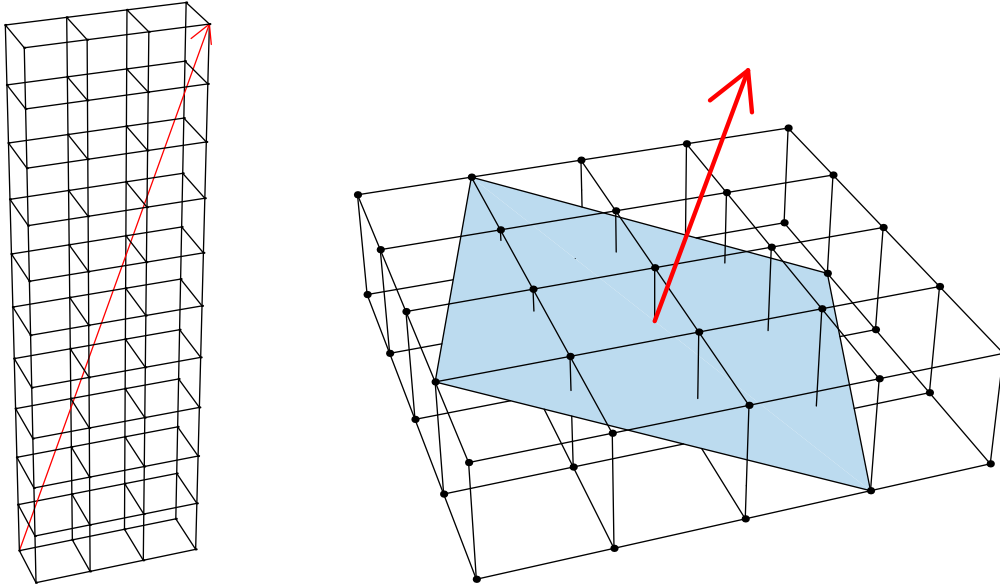
Figure 5.3: *Schematic view of the orientation of the line-of-sight (red line) and the lens planes (blue area) relative to the simulation box (indicated by black lines; figure by S. Hilbert).*

To avoid these problems, we exploit the periodic boundary conditions of the N-body simulation by arranging replicas of the simulation volume in a cubic lattice (Fig. 5.3). We then choose the l.o.s. at a skewed angle relative to the box axes. For certain choices of the direction of the l.o.s., the length of periodicity of the matter distribution along the l.o.s. can be made sufficiently large. At the same time, the matter distribution is continuous across the slice boundaries, preserving structures that are larger than the slices. On the other hand, we still wish to be able to apply Fourier methods for the calculation of the light deflection at the lens planes. This requires a matter density that is periodic perpendicular to the l.o.s. For an arbitrary choice of the l.o.s., this is clearly not the case. However, if one chooses the angles enclosed with the coordinate axes carefully, the lens planes can be made periodic again, even though with an area of periodicity that is in general much larger than the original $500 \times 500 \, h^{-2} \, \mathrm{Mpc}^2$.

To see this, we choose the origin of our coordinate system to be located at one of the box corners in the lattice of replicas of the simulation volume (see Fig. 5.3). If we set $L = 1$ for the moment, the positions of other box corners (*lattice points*) are then given by *lattice vectors* of the form $\boldsymbol{p} = (p_1, p_2, p_3)^{\mathrm{t}}$, where $p_i \in \mathbb{Z}$. We now choose two linearly independent lattice vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ with $\boldsymbol{p} = (p_1, p_2, p_3)^{\mathrm{t}}$ and $\boldsymbol{q} = (q_1, q_2, q_3)^{\mathrm{t}}$. These two vectors span a plane which is perpendicular to the lattice vector $\boldsymbol{n}$ with $\boldsymbol{n} = (n_1, n_2, n_3)^{\mathrm{t}} = \boldsymbol{p} \times \boldsymbol{q}$, $n_i \in \mathbb{Z}$. Since $\boldsymbol{p}$ and $\boldsymbol{q}$ are lattice vectors, the plane is itself periodic along the directions of $\boldsymbol{p}$ and $\boldsymbol{q}$ with a length of periodicity of $|\boldsymbol{p}|$ and $|\boldsymbol{q}|$, respectively. The total area of the parallelogram constructed from $\boldsymbol{p}$ and $\boldsymbol{q}$ is $|\boldsymbol{p} \times \boldsymbol{q}| = |\boldsymbol{n}|$. We show in App. B that there is no set of vectors spanning the same plane, but yielding a smaller area of periodicity, if the integer coefficients n_1, n_2, and n_3 are coprime. This also means that there is no shorter non-zero lattice vector perpendicular to the plane than $\boldsymbol{n}$ in this case, and hence, the shortest periodicity along the normal direction is $|\boldsymbol{n}|$.

Our choice $\boldsymbol{p} = (3, -1, 0)^{\mathrm{t}}$ and $\boldsymbol{q} = (1, 3, -1)^{\mathrm{t}}$ yields for the direction of the l.o.s. $\boldsymbol{n} = (1, 3, 10)^{\mathrm{t}}$, so that in physical units $|\boldsymbol{n}| = 5.244 \, h^{-1} \, \mathrm{Gpc}$. With this, the lens planes are rectangular with

(a) $z=z_k$:   (b) $z=z_{k+1}$:
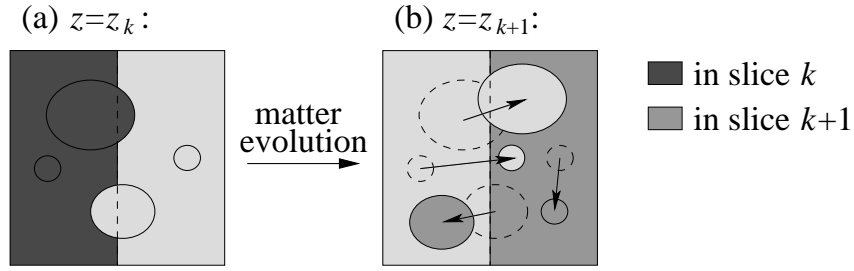

in slice k
in slice $k+1$

matter
evolution

Figure 5.4: *Schematic view of the adaptive slice boundaries to avoid the truncation or double inclusion of halos that are located near a slice boundary. Halos near the boundary of slice k and $k+1$ are either included as a whole in slice k or completely excluded depending on the positions of their centres (a). Halos that are included (excluded) in slice k, are excluded (included) from slice $k+1$ even if they have crossed the slice boundary between redshift k and $k+1$ (b) (figure by S. Hilbert).*

an area of periodicity of $1.581\,h^{-1}\,\mathrm{Gpc} \times 1.658\,h^{-1}\,\mathrm{Gpc}$. Moreover, any directions with shorter periodicity are at least $1.81°$ away from $\boldsymbol{n}$, and a light cone with a $1.7° \times 1.7°$ field of view does not intersect with itself up to redshift $z = 3.87$ when folded back into the simulation cube. The resulting orientation of the l.o.s and the lens planes with respect to the simulation box are illustrated in Fig. 5.3.

We choose the redshifts of the lens planes to be the same as the redshifts of the available snapshots for the Millennium Simulation. The slice boundaries are chosen to be located at the average redshifts between two lens planes. The boundary between two redshift slices with snapshot redshifts $z^{(k)}$ and $z^{(k+1)}$ is thus a plane at comoving distance $w_\mathrm{U}^{(k)} = w_\mathrm{L}^{(k+1)} = w\left[\left(z^{(k)} + z^{(k+1)}\right)/2\right]$ (with $w_\mathrm{L}^{(0)} = 0$). This means that in the special case of the slice for $z = 0$, we project the matter distribution onto a lens plane at $z^{(0)} = z^{(1)}/2 = 0.01$.

Special care is taken for the particles near the boundary of two slices. In order to avoid that a dark matter halo is only partially included into a slice (and hence would be only partially projected onto a lens plane), a halo is either included as a whole if its central particle is inside the slice as defined by boundary planes, or completely excluded otherwise. A complication may arise if a halo has moved across a slice boundary between two snapshots. In this case, we proceed as follows: if the halo is included into the slice of the snapshot with redshift z_1, its progenitors in the snapshot with redshift $z_2 > z_1$ are excluded from the slice for z_2 even if their central particles lie on the z_2-side of the slice boundary. This procedure is illustrated in Fig. 5.4.

As in the standard implementation, the lens planes are constructed from the matter slices by projection of the particle distribution parallel to the l.o.s. onto a rectangular mesh. In dense regions, the spatial resolution of the Millennium Simulation is effectively determined by the force softening, which is $\epsilon = 5\,h^{-1}\,\mathrm{kpc}$ (see Tab. 4.1). Thus, a mesh spacing of $2.5\,h^{-1}\,\mathrm{kpc}$ for the lens planes is required to avoid a degradation of the resolution for the projected matter density. However, a single mesh covering the full periodic area of the lens plane (i.e. $1.58\,\mathrm{Gpc}/h \times 1.66\,\mathrm{Gpc}/h$ comoving) with such a small mesh spacing would be computationally too demanding. We therefore adopt a strategy using a low resolution mesh covering the whole lens plane and several high-resolution meshes that can be computed on demand. Similar to the force calculation in the P^{3}M algorithm (Sec. 4.2), we split the lensing potential ψ on the lens planes into a long-range part ψ_l and a short-range part ψ_s. The split is defined in Fourier space by (see also

Eq. 4.19):

$$\tilde{\psi}_l(\boldsymbol{\ell}) = \tilde{\psi}(\boldsymbol{\ell}) \exp\left(-\theta_{\rm split}^2 \boldsymbol{\ell}^2\right) \; , \tag{5.19}$$

$$\tilde{\psi}_s(\boldsymbol{\ell}) = \tilde{\psi}(\boldsymbol{\ell}) \left[1 - \exp\left(-\theta_{\rm split}^2 \boldsymbol{\ell}^2\right)\right] \; . \tag{5.20}$$

Here $\theta_{\rm split} = x_{\rm split}/f_K^{(k)}$, where $x_{\rm split}$ is the comoving splitting length. We first assign the particles in a slice to the coarse mesh ($16\,384 \times 16\,384$ mesh points) using the CIC assignment scheme (see Sec. 4.1.2). We then use Eqs. (5.5) and (5.19) to compute the long-range potential ψ_l using FFT. The splitting length $x_{\rm split} = 0.175\,{\rm Mpc}/h$ is chosen to be slightly larger than the coarse mesh spacing ($0.096\,h^{-1}\,{\rm Mpc}$ and $0.101\,h^{-1}\,{\rm Mpc}$ in the directions given by $\boldsymbol{p}$ and $\boldsymbol{q}$, respectively), so that the coarse mesh samples $\psi_{\rm long}$ with sufficient accuracy. This is done only once for each lens plane, and the results are stored to disk for later use.

The short-range potential ψ_s is calculated during the actual ray-tracing only for that part of the lens plane which is intersected by light rays. If this area is larger than $40\,h^{-1}\,{\rm Mpc}$, it is subdivided into several sub-meshes. Each sub-mesh has a mesh spacing of $2.5\,h^{-1}\,{\rm kpc}$ and up to $16\,384 \times 16\,384$ mesh points. The fine meshes have to be slightly larger than the area defined by the light rays in order to take into account all matter within the effective range of ψ_s, which is given by $\approx 5x_{\rm split}$. Due to these extra margins, the potential can be computed for the region of interest using the FFT, even though the matter distribution in that patch is not periodic. This strategy leads to a significant speed-up in particular for low-redshift planes, where the comoving cross-section of the bundle of light rays is much smaller than the lens planes.

In order to reduce the shot noise from the individual particles, we employ two different smoothing schemes to construct the fine meshes. For the first, the particles are assigned to the fine mesh and subsequently smoothed with a Gaussian kernel of fixed comoving size. The second method is an adaptive, SPH-like smoothing scheme. Each particle contributes

$$\Sigma_p(\boldsymbol{x}) = \begin{cases} \dfrac{3m_p}{\pi r_p^2} \left(1 - \dfrac{|\boldsymbol{x} - \boldsymbol{x}_p|^2}{r_p^2}\right)^2 & \text{if } |\boldsymbol{x} - \boldsymbol{x}_p| < r_p, \\ 0 & \text{if } |\boldsymbol{x} - \boldsymbol{x}_p| \geq r_p, \end{cases} \tag{5.21}$$

to the surface mass density on the fine mesh. Here, $\boldsymbol{x}$ denotes comoving position on the lens plane, $\boldsymbol{x}_p$ is the projected comoving particle position, and r_p denotes the comoving distance to the 64th nearest neighbour particle in *three* dimensions (i.e. before projection). The adaptive smoothing is essentially equivalent to the assumption that, in three-dimensional space, each simulation particle represents a spherical cloud with a Gaussian density profile and an *rms* radius that is half the distance to its 64th nearest neighbour. This agrees with the notion that each particle actually represent a phase-space cell (see Chapter 4), the size of which is inversely proportional to the local phase-space density. The Gaussian smoothing scheme serves mostly for testing purposes; we choose the adaptive smoothing for all later applications.

Finally, the long- and short-range contributions to the deflection angles and shear matrices are calculated on the coarse and fine mesh by finite differencing of the potentials. The values between mesh points are obtained by bi-linear interpolation.

5.2 Applications

5.2.1 The convergence power spectrum

The convergence power spectrum is a central quantity in the analysis of cosmic shear measurements: on the one hand it can be computed directly from the three-dimensional matter
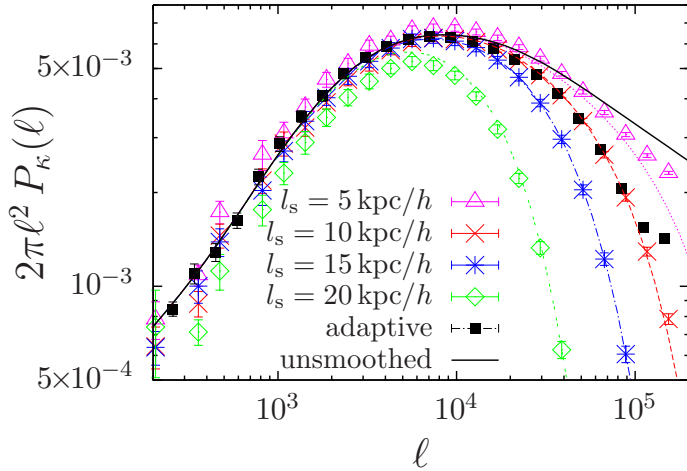
Figure 5.5: *Convergence power spectra for sources at $z = 1$ and with Gaussian smoothing (smoothing length l_s) on the lens planes (symbols with errorbars), compared to the prediction (dotted/dashed lines) obtained by projection of the measured power spectra of the actual mass distribution. The power spectrum from the runs using adaptive smoothing is given by the filled squares. The spectrum for the adaptive smoothing has been shifted slightly to the left for better visibility. The error bars were estimated from the field-to-field variance.*

power spectrum, for which theoretical models and fitting formulae are available. On the other hand, any commonly used two-point statistic can be computed from it. It is therefore of great interest to check the accuracy of the theoretical predictions for P_κ against ray-tracing simulations. Furthermore, the power spectrum constitutes a good tool to check the ray-tracing code as a whole.

If no ray-tracing simulation is at hand, P_κ is usually computed using Eq. (3.81), where for the three-dimensional matter power spectrum the fit formulae by Peacock & Dodds (1996) or Smith et al. (2003) are used. These, however, are based on N-body simulations that are comparable to the Millennium Simulation neither in box size nor in small-scale resolution. Furthermore, their accuracy even when compared the the original simulations is only of the order of a few percent. Therefore, this approach is not the optimal solution if we wish to explore the limits of our ray-tracing code. Instead, we make use of the three-dimensional matter power spectrum computed directly from the mass distribution in the Millennium Simulation. This guarantees that possible deviations of the simulated power spectra from the prediction are caused either by the failure of the Born approximation or by problems with our ray-tracing code. We use a discretization of Eq. (3.81) to obtain a prediction for P_κ (see also Vale & White 2003):

$$\hat{P}_\kappa(\ell) = \left(\frac{3H_0^2\Omega_m}{2c^2}\right)^2 \sum_{k=1}^{k_{max}} \left(w_U^{(k)} - w_L^{(k)}\right) \frac{g^2(w^{(k)})}{a^2(w^{(k)})} \hat{P}_\delta\left(\frac{\ell}{f_K(w^{(k)})}; w^{(k)}\right) , \qquad (5.22)$$

where $w^{(k)}$ is the comoving distance corresponding to the redshift of the k-th lens plane and $\hat{P}_\delta(k; w^{(k)})$ is the three-dimensional power spectrum measured from the k-th snapshot.

In Fig. 5.5, we present the results of several simulation sets with different smoothing schemes. In addition to the adaptive smoothing, which is intractable analytically, we have also employed Gaussian smoothing on the lens planes. The power spectrum measured from the adaptive smoothing runs shows excellent agreement with the prediction from Eq. (5.22) up to $\ell \approx 10^4$, where the smoothing becomes important. The simulations with Gaussian smoothing display a larger sampling variance (due to the smaller area of $80\,\text{deg}^2$ compared to $180\,\text{deg}^2$ for the simulations with adaptive smoothing). In this case, the error bars obtained from the field-to-field variances seem to underestimate the scatter due to cosmic variance in some cases; clearly, a larger set of simulations would be needed to improve the error estimates. This, however, does not affect the main conclusion from Fig. 5.5: in the case of Gaussian smoothing, which can be
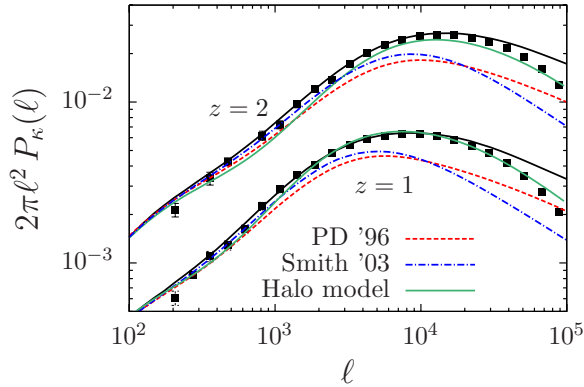
Figure 5.6: *Convergence power spectra for sources at z = 1 (lower curves) and z = 2 (upper curves). The simulation results are shown as symbols with errorbars, the corresponding internal predictions as solid lines. The predictions using the Peacock & Dodds (1996) prescription together with the transfer function from Eisenstein & Hu (1999) are given as dotted lines, those obtained from the Smith et al. (2003) fitting formula as dashed lines.*
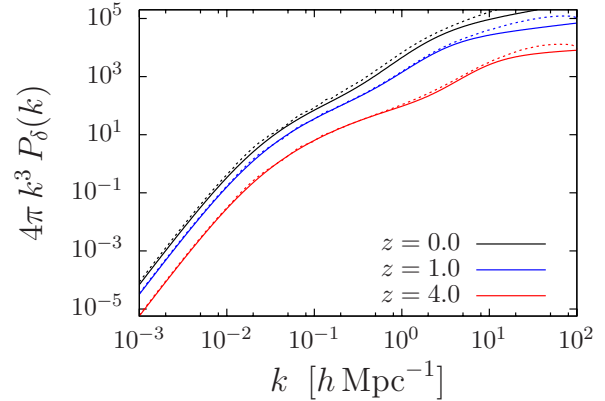
Figure 5.7: *Three-dimensional matter power spectra for z = 0.0, z = 1.0 and z = 4.0 (black, blue and red curves, respectively). The predictions made using the fit formula of Smith et al. (2003) are shown as solid curves, the shot-noise-corrected power spectra measured in the Millennium Simulation as dashed curves.*

modelled analytically, the power spectrum from the simulation is well understood on all scales and agrees with our prediction. Only the spectrum for the smallest smoothing length shows some aliasing effects on very small scales.

We note that it is not straightforward to make a statement about the small-scale resolution of our simulations. Even though the spectrum of the adaptive smoothing runs happens to match the spectrum for a Gaussian smoothing length of 10 kpc/h, this should not be interpreted as an "effective" smoothing length, because the smoothing scale of the adaptive scheme and therefore the small-scale resolution of the κ-maps is spatially variable. Furthermore, the resolution is a function of the source redshift distribution: if the mean source redshift is low, the lensing signal is created by low-redshift lens planes. Since the physical spacing of the light rays can become comparable to the spacing of FFT meshes on the first few planes, the finite resolution on the lens planes is more important than for sources at higher redshift. This is apparent from the fact that the measured power spectrum in Fig. 5.6 for sources at $z = 1$ deviates earlier from the prediction on small scales than for the sources at $z = 2$. Finally, also the prediction based on Eq. (5.22) is only of limited use in determining a scale beyond which the results of our simulations cannot be trusted, because the $\hat{P}_\delta(k; w^{(k)})$ were measured directly from the particle distribution in the simulation. As discussed before, this does not take into account the nature of the particles as tracers of phase-space elements of varying volume and therefore most likely overestimates the power on small scales. Therefore, a deviation of the measured power spectrum from Eq. (5.22) is not necessarily caused by numerical limitations and should not be used for defining the resolution of the ray-tracing simulation.

In Fig. 5.6 compare simulated power spectra for sources at $z = 1$ and $z = 2$ with predictions based on the popular fitting formulae by Peacock & Dodds (1996) (with the transfer function by Eisenstein & Hu 1999) and Smith et al. (2003). Both prescriptions strongly underpredict the power on small scales. This effect is already seen for the three-dimensional matter power
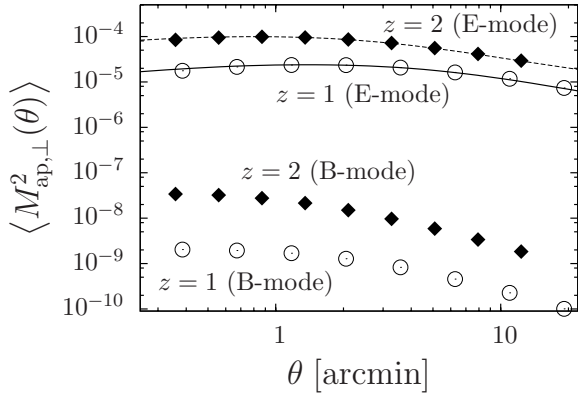
Figure 5.8: *Aperture mass dispersion for sources at $z = 1$ and $z = 2$. M_{E^2} from the simulation is given as symbols (E-modes: upper two sets of points, B-modes: lower sets), the semi-analytical prediction for the E-mode as solid lines. The error bars are much smaller than the symbols and are therefore omitted.*
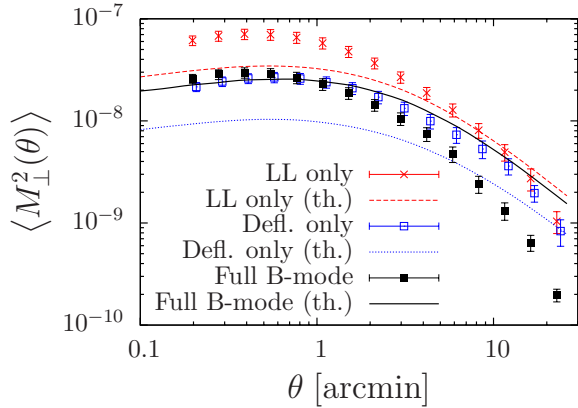
Figure 5.9: *B-modes as measured in the simulation, decomposed into the contributions by lens-lens-coupling and ray deflections. Shown as lines are the predictions based on Cooray & Hu (2002).*

spectra (see Fig. 5.7) and therefore seems to be a problem related to the fitting formulae for P_δ. Given their importance for the estimation of cosmological parameters from weak lensing surveys, a revision of the fitting formulae using more up-to-date N-body simulations would be highly desireable. We might even speculate that the lack of small-scale power in the predictions for P_δ is partly responsible for the discrepancy between weak lensing and other, independent methods (e.g. the analysis of the CMB), which seem to indicate a lower value of σ_8 than is found from lensing. Note that these considerations do not include the effects of baryonic physics on the power spectrum. Simulations indicate that this might increase the small-scale power even further (Rudd et al. 2008).

5.2.2 B-Modes

Finally, we wish to quantify the non-lensing systematic effects in our simulation. A suitable measure for this is the B-mode of the aperture mass dispersion $M_\perp^2$ (see Sec. 3.3.4).

As we have seen in Sec. 3.3.3, in the Born approximation cosmic shear does not create B-modes. This is no longer true in the full ray-tracing, where ray deflections are fully taken into account. We have given a theoretical prediction for the B-mode power spectrum that is accurate to the fourth order in the gravitational potential in (Eq. 3.84) by expanding the deflection angle. An equivalent approach was followed in Cooray & Hu (2002), who instead expanded the Jacobian matrix starting from the equation

$$\mathsf{A}_{ij}(\boldsymbol{\theta}, w) = \mathsf{E}_{ij} - \frac{2}{c^2} \int_0^w \mathrm{d}w' \, G(w, w') \, f_K(w') \, \Phi_{,ik}[\mathbf{x}(\boldsymbol{\theta}, w'), w'] \, \mathsf{A}_{kj}(\boldsymbol{\theta}, w') \,, \tag{5.23}$$

which can be obtained from Eq. (3.60) by differentiation with respect to $\boldsymbol{\theta}$. This allowed them to make a (slightly artificial) decomposition of the B-mode into a contribution from terms that are obtained by taking multiple deflections of the light rays into account (i.e. by expanding $\boldsymbol{x}$ in Eq. 5.23 using Eq. 3.59) and a contribution that arises from expanding A on the r.h.s. in

Eq. (5.23) in powers of Φ (which yields coupling terms of the potentials at different w, hence *lens-lens coupling*). Since the derivation of the B-mode power spectra due to these two effects involves basically the same steps and approximations as the derivation of the full B-mode that we have given in Sec. 3.3.3, we will not repeat it here and refer the reader to Cooray & Hu (2002) instead.

Fig. 5.8 shows the aperture mass dispersion measured from our set of simulations. As for the convergence power spectrum, we find excellent agreement of our prediction for the E-mode, based on the actual matter distribution in the Millennium Simulation, with the ray-tracing results. As expected from our theoretical considerations in Sec. 3.3.3, we also detect a significant B-mode. It is at least three orders of magnitude smaller than the E-mode, on larger scales their ratio even drops to more than 10^{-5}.

To disentangle possible numerical artifacts and true B-modes from physical higher-order effects, we switch off ray-deflections (i.e. we set $\mathsf{A}^{(k-1)} = \mathsf{E}$ in the second term on the l.h.s. of Eq. 5.13) or lens-lens-coupling (i.e. we set $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta} \; \forall k$ in Eq. 5.12) to determine the contributions of the individual effects to the total B-mode signal in Fig. 5.8.

The result is shown in Fig. 5.9, together with the theoretical predictions from Cooray & Hu (2002), for which we have also used the measured three-dimensional power spectra of the Millennium Run. Although the predictions are of the correct order of magnitude and reproduce some qualitative features of the simulations, the match is far from being perfect. While the B-mode on small scales is underpredicted by a factor of ≈ 2, the measured signal declines much more quickly on larger scales.

We have used both the polar decomposition (Eq. 3.33) and the spin decomposition (Eq. 3.38) of the Jacobian matrix to compute the shear field. We find that this does not lead to significant changes of the B-mode. Furthermore, we have excluded numerical effects in the estimation procedures as the reason for this discrepancy by applying different ways of estimating $M_\perp^2$ in real and Fourier space, as well as different methods of numerical integration for the theoretical predictions. To rule out a significant contribution of numerical artifacts from the ray-tracing code itself, we have run another set of simulations where both ray deflections and lens-lens-coupling were switched off. Only a very small B-mode (at least 6 orders of magnitudes smaller than the E-mode) remains. The origin of this tiny signal is found to be the interpolation of the shear matrices (see Eq. 5.13) from the meshes on the lens planes onto the light ray positions. This can be understood as follows: the sampling of an originally B-mode free, continuous shear field onto a grid and subsequent interpolation between the grid points again yields a continuous shear field. This, however, agrees with the original field exactly only at the grid points. Therefore, it will in general contain a small B-mode contribution, depending on the grid resolution and the interpolation scheme used.

The origin of the discrepancy therefore remains unclear. Possible reasons are the quality of some of the approximations made in the derivation of the B-mode power spectrum, or the fact that we have computed the shear field from the simulations using the form of the Jacobian as defined in Eq. (3.33), whereas for the theoretical prediction the definition in Eq. (3.38) was used. These two agree in the limit of weak lensing ($\kappa \ll 1$), which, however, is not fulfilled everywhere in the simulations.

Nevertheless, the relatively small difference between theory and simulation does not affect the main conclusion of Shapiro & Cooray (2006): the cosmic shear B-mode signal is undetectable for any realistic weak lensing survey. Therefore, it is a valid approach to use $M_\perp^2$ as an indicator for systematic effects resulting from the data reduction when working with real weak lensing data.

6 Weak Lensing and Statistics

Statistical methods play a very important role in any weak lensing study: the detection of the weak distortion patterns can only be achieved by looking at large samples of background galaxies; furthermore, the underlying shear and density fields can only be described by their statistical properties. Finally, the methods of statistical inference are needed to extract information about cosmological parameters from the data and to quantify their uncertainties. In this chapter, we discuss several aspects related to the inference problem. We first review the basic theory of Bayesian inference, followed by a discussion of how to obtain good estimates of the inverse covariance matrix and therefore of the errors on inferred parameters. We then investigate the accuracy of approximating the cosmic shear likelihood function by a Gaussian and provide a method to estimate the likelihood function from simulations. Finally, we apply our results to cosmic shear measurements in the Chandra Deep Field South (CDFS), for which in Schrabback et al. (2007) an extraordinarily low power spectrum normalization σ_8 has been found. We investigate what role the non-Gaussianity of the likelihood might play in this, and what share cosmic variance and the criteria according to which the CDFS was selected have in this matter.

6.1 Statistical inference

Until today, statisticians have not agreed on a common definition of probability. The field of statistics is therefore divided into two schools: orthodox statisticians ("frequentists") define probability strictly as the frequency of the occurrence of an event, whereas in the Bayesian paradigm probability quantifies the plausibility of a proposition. A discussion of the relative merits and shortcomings of either point of view is considerably beyond the scope of this thesis (but see, e.g., Jaynes 2003, 1983; MacKay 2003; Loredo 1989), and since cosmologists mostly adhere to the latter school of thought, we will perform our analysis within the Bayesian paradigm (for more details, see e.g. Gelman et al. 2004; Jaynes 2003; MacKay 2003).

6.1.1 Parameter estimation

Let us assume that we have measured some data $\boldsymbol{d} \in \mathbb{R}^p$ and now wish to infer the parameters $\boldsymbol{\pi}$ of a model that provides us with a prediction $\boldsymbol{m}(\boldsymbol{\pi})$ for the data. The model also has to provide the mechanism by which the data are generated, i.e. it includes the probability distribution $p(\boldsymbol{d}|\boldsymbol{\pi})$ of the data given a certain set of model parameters. This probability distribution is called the *likelihood*. Adopting a Bayesian point of view, our aim is to compute the *posterior likelihood*, i.e. the probability distribution $p(\boldsymbol{\pi}|\boldsymbol{d})$ of the parameter vector $\boldsymbol{\pi}$ given the information provided by the data $\boldsymbol{d}$. We can relate the posterior to the likelihood using Bayes' theorem:

$$p(\boldsymbol{\pi}|\boldsymbol{d}) = \frac{p(\boldsymbol{\pi})}{p(\boldsymbol{d})} p(\boldsymbol{d}|\boldsymbol{\pi}) . \tag{6.1}$$

Here, $p(\boldsymbol{\pi})$ is the prior distribution of the parameters, which incorporates our knowledge about $\boldsymbol{\pi}$ prior to looking at the data; such can originate from previous measurements or theoretical

arguments. Several methods exists even to express total ignorance about the value of $\boldsymbol{\pi}$, the simplest being to choose a uniform prior $p(\boldsymbol{\pi}) = \text{const}$. The evidence $p(\boldsymbol{d})$ in this context simply serves as a normalization factor, but becomes important for Bayesian model comparison.

In the field of weak gravitational lensing, it is almost universally assumed that the likelihood $p(\boldsymbol{\xi}|\boldsymbol{\pi})$ is a Gaussian distribution:

$$p(\boldsymbol{d}|\boldsymbol{\pi}) \propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\pi})\right]^{\mathrm{t}} \mathsf{C}^{-1}(\boldsymbol{\pi})\left[\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\pi})\right]\right\} \quad , \tag{6.2}$$

where $\mathsf{C}(\boldsymbol{\pi})$ is the covariance matrix of $\boldsymbol{d}$ as predicted by the underlying model. Usually, however, the dependence of the covariance matrix upon cosmological parameters is not taken into account. Rather, the covariance that is computed for a fixed fiducial set of parameters $\boldsymbol{\pi}_0$ is used in Eq. (6.2). Under this approximation, the likelihood is a function of the difference $\boldsymbol{\Delta}(\boldsymbol{\pi}) = \boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\pi})$ only:

$$p(\boldsymbol{d}|\boldsymbol{\pi}) = L_{\boldsymbol{\pi}_0}\left[\boldsymbol{\Delta}(\boldsymbol{\pi})\right] \quad . \tag{6.3}$$

Sometimes, one is only interested in a subset $\boldsymbol{\pi}_1$ of $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)^{\mathrm{t}}$. This is the case for example if the parameters in $\boldsymbol{\pi}_2$ are so-called *nuisance parameters* (e.g. parametrizing systematic effects) or if one wishes to reduce the dimensionality of the parameter space for purposes of visualization of the posterior density. The "uninteresting" parameters can be eliminated by *marginalization*:

$$p(\boldsymbol{\pi}_1|\boldsymbol{d}) = \int \mathrm{d}\boldsymbol{\pi}_2\, p(\boldsymbol{\pi}|\boldsymbol{d}) \,. \tag{6.4}$$

Note that $p(\boldsymbol{\pi}_1|\boldsymbol{d})$ is different from the posterior that would have been obtained by only fitting for $\boldsymbol{\pi}_1$ and keeping $\boldsymbol{\pi}_2$ fixed (i.e. using $\boldsymbol{\pi}_1$ in Eq. 6.1 instead of $\boldsymbol{\pi}$): the marginalized posterior in Eq. (6.4) still accounts for the uncertainties of the nuisance parameters.

All information about $\boldsymbol{\pi}$ provided by the data is contained in the posterior distribution. It is, however, customary to quote "best-fit" parameters and corresponding error estimates. Such point estimates of the parameters can be obtained from the posterior in several ways, minimizing different *risk functions*. The risk of a point estimate $\hat{\boldsymbol{\pi}}$ is the expected value of a *loss function*, which quantifies the "loss" of making a wrong estimate. Common loss functions are the quadratic loss, $L_2(\hat{\boldsymbol{\pi}}, \boldsymbol{\pi}_0) = (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0)^2$, where $\boldsymbol{\pi}_0$ is the true parameter vector, the L_1-loss $L_1(\hat{\boldsymbol{\pi}}, \boldsymbol{\pi}_0) = |\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0|$ and the $0 - 1$–loss $L_{01}(\hat{\boldsymbol{\pi}}, \boldsymbol{\pi}_0) = 1$ if $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}_0$ and $= 0$ else. In Bayesian statistics, the risk function $R(\hat{\boldsymbol{\pi}})$ then is the expectation value of the loss over the posterior distribution:

$$R(\hat{\boldsymbol{\pi}}) = \int \mathrm{d}\boldsymbol{\pi}\, p(\boldsymbol{\pi}|\boldsymbol{d})\, L(\hat{\boldsymbol{\pi}}, \boldsymbol{\pi}) \,. \tag{6.5}$$

Minimizing the three risks of the three loss functions introduced above with respect to $\hat{\boldsymbol{\pi}}$, we find the following point estimates:

- The quadratic risk is minimized by the expectation value of the parameters, defined by

$$\hat{\boldsymbol{\pi}} = E(\boldsymbol{\pi}) = \int \mathrm{d}\boldsymbol{\pi}'\, \boldsymbol{\pi}'\, p(\boldsymbol{\pi}'|\boldsymbol{d}) \,. \tag{6.6}$$

- If $\boldsymbol{\pi} \in \mathbb{R}$, i.e. if only one parameter is considered, the median of the posterior minimizes the risk based on the loss function L_1.

- The *maximum a posteriori* (MAP) estimate is determined by the maximum of $p(\boldsymbol{\pi}|\boldsymbol{\xi})$ and minimizes the $0 - 1$-risk. Note that in the case of a uniform prior, this point estimate is equivalent to the maximum likelihood estimate.

In Bayesian inference, error estimates are given by so-called credible intervals/regions. An $\alpha\%$ credible region is defined as the region containing a fraction $\alpha/100$ of the probability around a point estimate. To make this definition unique, further conditions are needed. For the *highest-posterior-density* (HPD) intervals, those regions of the parameter space are used which contain the highest values of the posterior, including the mode. These intervals are the narrowest possible. Another possible choice in one dimension is to choose the interval such that the probability of the parameter value being larger than the upper interval boundary is as large as the probability for it to be below the lower limit.

6.2 Unbiased estimation of the inverse covariance matrix

In this section, we investigate possible biases that may arise if the covariance matrix for the Gaussian likelihood function (Eq. 6.2) is estimated from simulations or from the data instead of using a model prediction. We distinguish between the *population covariance* $\boldsymbol{\Sigma}$, which is the true covariance matrix of the data $\boldsymbol{d}$, and the estimated *sample covariance* matrix, which we denote by $\hat{\mathsf{C}}$. For notational convenience, we also introduce the log-likelihood

$$\mathcal{L}(\boldsymbol{d}|\boldsymbol{\pi}) \propto -\frac{1}{2} \left[\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\pi})\right]^{\mathrm{t}} \boldsymbol{\Sigma}^{-1} \left[\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\pi})\right] \ . \tag{6.7}$$

For the evaluation of the likelihood the population covariance matrix $\boldsymbol{\Sigma}$ and its inverse $\boldsymbol{\Sigma}^{-1}$ or estimates thereof are needed. In most cases, no exact analytical expression for $\boldsymbol{\Sigma}$ can be given, although numerous authors make use of analytical approximations. An example from the field of weak gravitational lensing is Semboloni et al. (2006), who use the Gaussian approximation to the covariance matrix of the shear correlation functions given by Schneider et al. (2002a). Other possibilities are to estimate $\boldsymbol{\Sigma}$ from the data themselves (e.g. Hetterscheidt et al. 2005; Budavári et al. 2003) or to obtain it from a simulated data set whose properties are comparable to the original data (e.g. Pan & Szapudi 2005). In the latter paper, the authors observed that the estimated covariance matrix becomes singular if p, the number of entries of the data vectors, exceeds the number of observations / simulated data vectors. As a remedy, they propose to use the Singular Value Decomposition (SVD, Press et al. 1992) to obtain a pseudo-inverse of the covariance matrix, but do not investigate the properties of the resulting estimate of $\boldsymbol{\Sigma}^{-1}$ in detail. In the following, we prove analytically that the rank of the standard estimator of the covariance matrix cannot exceed the number of observations. We then point out that, even if this estimator is not singular, simple matrix inversion yields a biased estimator of $\boldsymbol{\Sigma}^{-1}$. This may, if not corrected for, cause a serious underestimate of the size of the credible regions; a fact which – to our knowledge – has not been reported before in the astronomical literature. We report on the existence of a method to remove this bias, which can be derived for Gaussian noise and statistically independent data vectors, and test the validity of this method when these assumptions are violated.

6.2.1 The covariance matrix

Estimators

Let $\boldsymbol{d}$ be a vector of p random variables with components d_i, drawn from a multi-variate Gaussian distribution with population covariance matrix Σ and mean $\boldsymbol{\mu}$:

$$P(\boldsymbol{d}) = \frac{1}{(2\pi)^{p/2}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{\mu})^{\mathrm{t}}\Sigma^{-1}(\boldsymbol{d}-\boldsymbol{\mu})\right) . \tag{6.8}$$

Furthermore, let $\boldsymbol{d}^{(k)}$ denote the k-th realization of this random vector, where $k \in [1, N]$ and N is the total number of realizations. The well-known maximum-likelihood estimator for the components of the covariance matrix is given by (e.g. Barlow 1991)

$$\hat{\mathsf{C}}_{ij}^{\mathrm{ML}} = \frac{1}{N}\sum_{k=1}^{n} \left(d_i^{(k)}-\mu_i\right)\left(d_j^{(k)}-\mu_j\right) , \tag{6.9}$$

which in the case of a known mean vector $\boldsymbol{\mu}$ is unbiased. If, however, $\boldsymbol{\mu}$ has to be estimated from the data, a correction factor of $N/(N-1)$ has to be applied to Eq. (6.9).

The rank of $\hat{\mathsf{C}}^{\mathrm{ML}}$

In the following, we prove that $\hat{\mathsf{C}}^{\mathrm{ML}}$ is singular for $p > N$ in case of known mean vector, and for $p > N - 1$ if the mean vector is obtained from the data as well. For the first case, this can be seen by rewriting Eq. (6.9) as

$$\hat{\mathsf{C}}^{\mathrm{ML}} = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{d}^{(k)}\,\boldsymbol{d}^{(k)\,\mathrm{t}} , \tag{6.10}$$

where we presume, without loss of generality, that the mean vector is zero. Since the data vectors $\boldsymbol{d}^{(k)}$ are statistically independent, we can safely assume that they are linearly independent for $N \leq p$ (for a continuous distribution, the probability to draw linearly dependent data vectors is zero). Therefore, $\left\{\boldsymbol{d}^{(k)}\right\}$ span an N-dimensional subspace U of $\mathbb{R}^p$. To check whether $\hat{\mathsf{C}}^{\mathrm{ML}}$ is singular we now try to find a vector $\boldsymbol{y} \neq \boldsymbol{0}$ for which $\hat{\mathsf{C}}^{\mathrm{ML}}\,\boldsymbol{y} = \boldsymbol{0}$. Looking at Eq. (6.10), we see that this is only possible for $p > N$, since in this case we can always choose a vector $\boldsymbol{y}$ from the subspace orthogonal to U, for which $\boldsymbol{d}^{(k)} \cdot \boldsymbol{y} = 0 \,\forall\, k$. If $p \leq N$, $\left\{\boldsymbol{d}^{(k)}\right\}$ already spans the whole of $\mathbb{R}^p$, and no vector can be found that is orthogonal to all $\boldsymbol{d}^{(k)}$. This proves that $\hat{\mathsf{C}}^{\mathrm{ML}}$ is singular for known mean vector if $p > N$.

We now prove our statement for an unknown mean vector $\boldsymbol{\mu}$, which is estimated from the data using

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{d}^{(k)} . \tag{6.11}$$

For this, we define a new set of independent data vectors $\left\{\boldsymbol{w}^{(k)}\right\}$ by forming linear combinations of $\left\{\boldsymbol{d}^{(k)}\right\}$, specified by the orthogonal transformation B, of which we demand that the last

(N-th) row be given by $(1/\sqrt{N}, \dots, 1/\sqrt{N})$ (Anderson 2003):

$$\boldsymbol{w}^{(k)} = \sum_{l=1}^{N} \mathsf{B}_{kl} \, \boldsymbol{d}^{(l)} \ . \tag{6.12}$$

Thanks to our choice of B_{Nl}, we have $\boldsymbol{w}^{(N)} = \sqrt{N}\,\boldsymbol{\mu}$. Next, we rewrite $\hat{\mathsf{C}}^{\mathrm{ML}}$ by means of the new data vectors:

$$\hat{\mathsf{C}}^{\mathrm{ML}} \ = \ \frac{1}{N} \sum_{k=1}^{N} \boldsymbol{d}^{(k)} \, \boldsymbol{d}^{(k)\mathrm{t}} - \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{t}} \tag{6.13}$$

$$= \ \frac{1}{N} \sum_{k=1}^{N} \boldsymbol{w}^{(k)} \, \boldsymbol{w}^{(k)\mathrm{t}} - \frac{1}{N}\boldsymbol{w}^{(N)} \, \boldsymbol{w}^{(N)\mathrm{t}} \tag{6.14}$$

$$= \ \frac{1}{N} \sum_{k=1}^{N-1} \boldsymbol{w}^{(k)} \, \boldsymbol{w}^{(k)\mathrm{t}} \ . \tag{6.15}$$

The last expression is of the same form as Eq. (6.10) (except for the sum, which has one addend less), and so the same line of reasoning as above can be applied to show that $\hat{\mathsf{C}}^{\mathrm{ML}}$ is singular for $p > N - 1$.

Another interesting implication of Eq. (6.15) is that the mean vector and the estimated covariance matrix are distributed independently (again see Anderson 2003), although they are computed from the same data vectors. First, note that $\boldsymbol{w}^{(i)}$ and $\boldsymbol{w}^{(j)}$ are statistically independent for $i \neq j$. This can be seen by computing the covariance between the two vectors:

$$\mathrm{Cov}\left(\boldsymbol{w}^{(i)}, \boldsymbol{w}^{(j)}\right) \ = \ \left\langle \left(\boldsymbol{w}^{(i)} - \boldsymbol{\nu}^{(i)}\right) \left(\boldsymbol{w}^{(j)} - \boldsymbol{\nu}^{(j)}\right)^{\mathrm{t}} \right\rangle \tag{6.16}$$

$$= \ \sum_{k,l=1}^{N} \mathsf{B}_{ik}\mathsf{B}_{jl} \left\langle \left(\boldsymbol{d}^{(k)} - \boldsymbol{\mu}\right) \left(\boldsymbol{d}^{(l)} - \boldsymbol{\mu}\right)^{\mathrm{t}} \right\rangle \tag{6.17}$$

$$= \ \sum_{k,l=1}^{N} \mathsf{B}_{ik}\mathsf{B}_{jl} \, \delta_{kl} \, \Sigma \tag{6.18}$$

$$= \ \delta_{ij} \, \Sigma \tag{6.19}$$

Here, $\langle \cdot \rangle$ denotes the expectation value and $\boldsymbol{\nu}^{(i)} = \boldsymbol{\mu} \sum_{j=1}^{N} \mathsf{B}_{ij}$ is the mean value of $\boldsymbol{w}^{(i)}$. Since $\hat{\mathsf{C}}^{\mathrm{ML}}$ does not depend on $\boldsymbol{w}^{(N)}$, which in turn is statistically independent of the remaining $\boldsymbol{w}^{(i)}$, this shows the independence of estimated mean and covariance.

6.2.2 The inverse covariance matrix

An unbiased estimator for Σ^{-1}

From Eq. (6.9), an estimator for Σ^{-1} can be obtained by matrix inversion:

$$\hat{\mathsf{C}}_*^{-1} = \left(\hat{\mathsf{C}}^{\mathrm{ML}}\right)^{-1} \ . \tag{6.20}$$

This estimator is consistent, but *not unbiased* due to noise in $\hat{\mathsf{C}}^{\mathrm{ML}}$: the inverse of an unbiased estimator for some statistical variable X is in general not an unbiased estimator for X^{-1}. Indeed,
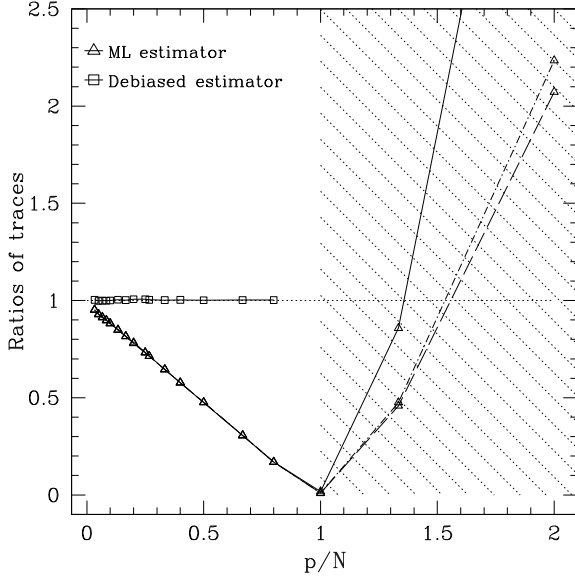
Figure 6.1: *Ratios of the trace of Σ^{-1} to the traces of C_*^{-1} (triangles) and $\hat{\mathsf{C}}^{-1}$ (squares), respectively. The dashed line is for the covariance model Eq. (6.23), the solid line for Eq. (6.24) and the dot-dashed-line for Eq. (6.25). The original data vectors had $p_1 = 240$ bins, and were re-binned by subsequently joining $2, 3, \ldots$ of the original bins. The number of independent observations is $N = 60$. Error bars are comparable to the symbol size and therefore omitted.*

in our case of Gaussian errors and statistically independent data vectors one can show (Anderson 2003) that the expectation value of $\hat{\mathsf{C}}_*^{-1}$ is *not* the inverse of the population covariance, but

$$\left\langle \hat{\mathsf{C}}_*^{-1} \right\rangle = \frac{\nu}{\nu - p - 1} \, \Sigma^{-1} \qquad \text{for } p < \nu - 1 \,, \tag{6.21}$$

where $\nu = N$ if $\boldsymbol{\mu}$ is known and $\nu = N - 1$ if the mean is estimated from the data. In the following, we will only pursue the latter case. The bias in $\hat{\mathsf{C}}_*^{-1}$ thus depends essentially on the ratio of the number of entries p in the data vectors (henceforth referred to as the number of bins) to the number of independent observations N. From Eq. (6.21) it follows that an unbiased estimator of Σ^{-1} is given by[1]

$$\hat{\mathsf{C}}^{-1} = \frac{N - p - 2}{N - 1} \, \hat{\mathsf{C}}_*^{-1} \qquad \text{for } p < N - 2 \,. \tag{6.22}$$

6.2.3 Monte-Carlo experiments

To illustrate Eq. (6.22), we perform the following experiment: First, we choose an analytical form for the population covariance Σ. We use three different models:

$$\Sigma_{ij}^{\mathrm{d,c}} = \sigma^2 \, \delta_{ij} \,, \tag{6.23}$$
$$\Sigma_{ij}^{\mathrm{d,l}} = \sigma^2 \, [1 - i/(1 + p_1)] \, \delta_{ij} \quad \text{and} \tag{6.24}$$
$$\Sigma_{ij}^{\mathrm{nd}} = \sigma^2/(1 + \epsilon|i - j|) \,, \tag{6.25}$$

which initially are $p_1 \times p_1$ matrices. ϵ can be used to tune the degree of correlation in model (6.25); we choose $\epsilon = 0.05$.

---

[1]Note that there is a typing error in Anderson's book, where he gives an expression corresponding to

$$\hat{\mathsf{C}}^{-1} = \frac{N - p - 2}{N - \mathbf{2}} \, \hat{\mathsf{C}}_*^{-1} \text{ for } p < N - 2.$$

We then create N data vectors of length p_1 according to $\boldsymbol{d}^{(k)} = \boldsymbol{m} + \boldsymbol{\gamma}^{(k)}(\Sigma_1)$, where $\boldsymbol{\gamma}^{(k)}(\Sigma_1)$ is a noise vector drawn from a multivariate Gaussian distribution with mean zero and covariance Σ_1. The choice of the model vector $\boldsymbol{m}$ is arbitrary, and in fact for the present purpose it would be sufficient to set $\boldsymbol{m} = \boldsymbol{0}$. For later use, however, we choose the linear model $m_i = ax_i + b$, where $x_i = (x_{\max} - x_{\min})(i + 1/2)/p_1$ is the value of the free variable corresponding to the centre of the i-th bin.

The noise vectors are constructed by means of the Cholesky decomposition (Press et al. 1992) of Σ, which yields a lower triangular matrix T such that $\Sigma = \mathsf{T}\mathsf{T}^{\mathrm{t}}$. Noise vectors with the desired distribution can then be obtained by $\boldsymbol{\gamma}^{(k)} = \mathsf{T}\,\boldsymbol{r}^{(k)}$ (Hu & Keeton 2002), where $\boldsymbol{r}^{(k)}$ is a vector of independent random numbers drawn from the normal distribution.

From this synthetic set of observations we estimate the mean data vector and the covariance matrix, which yields the estimator $\hat{\mathsf{C}}^{\mathrm{ML}}$. Next, both Σ and $\hat{\mathsf{C}}^{\mathrm{ML}}$ are inverted using the Singular Value Decomposition (see below). Finally, we compute the unbiased estimate $\hat{\mathsf{C}}^{-1}$ of the inverse covariance as given in Eq. (6.22).

To probe the dependence of the bias of the estimators for Σ^{-1} on the number of bins, we create samples of N new data vectors each for several values of p. For that, we assume that the new data vectors, denoted by $^{(j)}\boldsymbol{d}$, with $p_j = p_1/j$ ($j = 2 \ldots p_1/2$) bins are obtained by averaging over j bins of the original data vector $^{(1)}\boldsymbol{d}$:

$$^{(j)}d_i = \frac{1}{j} \sum_{k=0}^{j-1} {}^{(1)}d_{j \cdot i - k} \ . \tag{6.26}$$

The population covariance $^{(j)}\Sigma$ for a data vector with p_j bins is obtained from the original $^{(1)}\Sigma$ by averaging over $(j \times j)$-sub-blocks of $^{(1)}\Sigma$:

$$^{(j)}\Sigma_{ab} = \frac{1}{j^2} \sum_{k=0}^{j-1} \sum_{l=0}^{j-1} {}^{(1)}\Sigma_{ja-k,jb-l} \ . \tag{6.27}$$

This strategy of re-binning has the advantage that the true covariance is known exactly for all p_j. Strictly speaking, this way of re-binning is only optimal if the errors in the bins which are joined are uncorrelated and equal. If this is not the case, one should perform a suitably weighted average over the sub-blocks, which would slow down the simulations considerably. However, testing has shown that this does not change our results significantly, which is why we use unweighted averaging also for the covariance models (6.24) and (6.25). Furthermore, note that in practice we do not re-bin the original data vectors, but create new noise realizations for each binning to avoid correlations between data points for different p.

Since the bias in Eq. (6.21) is just a scalar factor, we record the traces of the estimators $\hat{\mathsf{C}}_*^{-1}$ and $\hat{\mathsf{C}}^{-1}$ for each number of bins p. To improve our statistics, we repeat the procedure outlined above 10^4 times and average over the traces computed in each step.

In Fig. 6.1, we plot the ratios of the trace of Σ^{-1} to the traces of $\hat{\mathsf{C}}_*^{-1}$ and $\hat{\mathsf{C}}^{-1}$, respectively. Not using the bias-corrected $\hat{\mathsf{C}}^{-1}$ can have considerable impact on the size of confidence regions of parameter estimates: for $p < N - 2$, the components of $\hat{\mathsf{C}}_*^{-1}$ will be too large compared to those of the true inverse covariance, and the log-likelihood will decrease too steeply. This leads to confidence contours too small; the errors of the parameter estimates are underestimated the more, the closer one gets to $p = N - 1$. As expected, for $p < N - 2$ the bias of $\hat{\mathsf{C}}_*^{-1}$ does not depend on the covariance model.

We also plot the traces of $\hat{\mathsf{C}}_*^{-1}$ for the different covariance models beyond $p \geq N - 1$, where the estimator $\hat{\mathsf{C}}^{\mathrm{ML}}$ is singular. These data points have been obtained using the Singular Value Decomposition to invert the covariance matrix, yielding a decomposition of the form

$$\mathsf{C} = \mathsf{U}\mathsf{W}\mathsf{V}^{\mathrm{t}} \,, \tag{6.28}$$

where U and V are orthogonal matrices and W is a diagonal matrix containing the singular values. Since C is symmetric, one has in addition $\mathsf{U} = \mathsf{V}$, while W contains the absolute values of the eigenvalues of C. The inverse of C is then given by $\mathsf{C}^{-1} = \mathsf{V}\mathsf{W}^{-1}\mathsf{U}^{-1}$. If C is singular, some of the entries of W will be zero or comparable to machine precision. We therefore can only compute a pseudo-inverse of C by replacing the inverses of these singular values in W^{-1} by zero, as has been suggested in Press et al. (1992) and Pan & Szapudi (2005). Fig. 6.1 shows that the bias of $\hat{\mathsf{C}}_*^{-1}$ in this regime depends significantly on the covariance model chosen and does not depend on binning in a simple way. Therefore we strongly discourage from the use of the SVD for $p > N - 1$.

6.2.4 Implications for likelihood analysis

Bias of the log-likelihood function

Having obtained an unbiased estimator of the inverse covariance matrix, one may still be concerned about a possible bias in the log-likelihood function, since it consists of the product of $(\boldsymbol{d} - \boldsymbol{\mu})$ and $\hat{\mathsf{C}}^{-1}$ (Eq. 6.7), both of which are estimated from the same set of observations. In other words, the question is if it is possible to write

$$
\begin{aligned}
\langle \mathcal{L}(\boldsymbol{d}|\boldsymbol{\pi}) \rangle &= -\frac{1}{2} \left\langle (\boldsymbol{\mu} - \boldsymbol{m})^{\mathrm{t}} \, \hat{\mathsf{C}}^{-1} \, (\boldsymbol{\mu} - \boldsymbol{m}) \right\rangle \\
&= -\frac{1}{2} \left\langle (\boldsymbol{\mu} - \boldsymbol{m})^{\mathrm{t}} \right\rangle \left\langle \hat{\mathsf{C}}^{-1} \right\rangle \left\langle (\boldsymbol{\mu} - \boldsymbol{m}) \right\rangle \,.
\end{aligned}
\tag{6.29}
$$

This is indeed the case, since we have shown at the end of Sect. 6.2.1 that mean vector and covariance matrix are distributed independently.

Again, we verify this result using our simulation. During the course of the procedure given above, we also compute the log-likelihood function for the straight-line fit on a grid in parameter space, using both $\hat{\mathsf{C}}^{-1}$ and $\hat{\mathsf{C}}_*^{-1}$. To detect a possible bias without introducing an additional one through the measurement, we need a function which is linear in the log-likelihood. An obvious choice is the sum over all grid cells of the log-likelihood grid. In Fig. 6.2, we show the ratio of the likelihood sum, computed using the biased and unbiased estimators of the covariance matrix, to the true likelihood sum. Consistent with what we said in the preceding section, the estimator $\hat{\mathsf{C}}_*^{-1}$ leads to a likelihood function which is too steep, and therefore to the underestimation of the size of the confidence regions. In contrast, no such effect is present if one uses $\hat{\mathsf{C}}^{-1}$ instead.

Marginalized likelihood

If we marginalize over all but one model parameter, the marginalized log-likelihood is given by

$$\mathcal{L}_i(\boldsymbol{d}|\pi_i) = \log \left\{ \left[\prod_{j \neq i} \int \mathrm{d}\pi_j \right] \exp\left[\mathcal{L}(\boldsymbol{d}|\boldsymbol{\pi}) \right] \right\} \,. \tag{6.30}$$
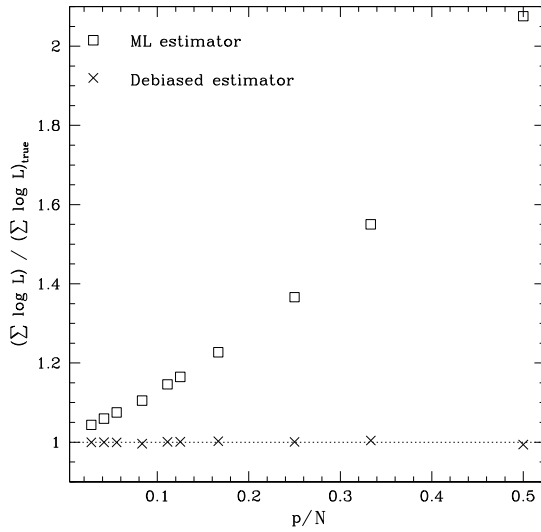
Figure 6.2: *Ratios of the sums over all grid points of the log-likelihood functions computed using the inverted ML-estimator Eq. (6.20) (open squares) and the de-biased estimator Eq. (6.22) (crosses) to the sum over all grid points of the true log-likelihood.*
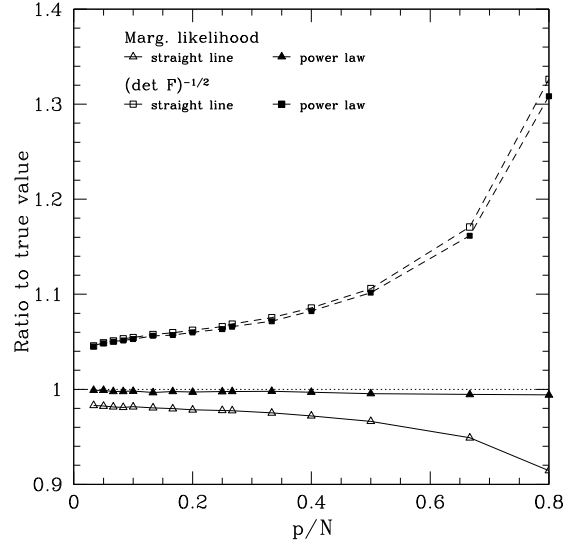
Figure 6.3: Triangles, solid lines: *Ratio of the sum over all pixels of the marginalized likelihood computed using $\hat{C}^{-1}$ to the one obtained using the true likelihood. Filled triangles are for the power-law fit (marginalized over the power law index), open triangles are for the straight line fit (marginalized over the intercept). Squares, dashed lines: Ratio of $\sqrt{\det F^{-1}}$ using $\hat{C}^{-1}$ to the true one, computed with Σ. For both cases, $\Sigma = \Sigma^{d,c}$ (Eq. 6.23).*

There is no reason to believe that the marginalized log-likelihood, and with it the size of the errors on π_i bars, are unbiased, even if one uses the unbiased $\hat{C}^{-1}$. We demonstrate this by means of our simulated fitting procedure, where we now use not only the straight line model, but perform a second simulation using a power law model of the form $m_i = a\,x_i^b$. We marginalize over the intercept of the line and the power law index, respectively. Similar to what we did in the case of the full parameter space, we record the sums over all pixels of the (one-dimensional) grid of the marginalized log-likelihood functions, which we compute using the true Σ^{-1} and the estimator $\hat{C}^{-1}$. For Σ, we choose the model (6.23). We average over $\approx 3 \times 10^4$ repetitions of these experiments. We plot the ratio of true to estimated log-likelihood sums in Fig. 6.3 (triangles and solid lines). The plot shows a bias of maximally $\approx 8\%$ for the straight line and even less for the power law, in a direction which would lead to an overestimation of the error bars on slope and amplitude. Although the effect is not very large, this is not guaranteed to remain so for models different from the ones considered here.

Measuring the size of credible regions

For some applications, it is useful to have a simple measure of the size of the credible regions. As an example, we make use of the Fisher information matrix F (Fisher 1935), which is defined

by (e.g. Kendall et al. 1987)

$$\mathsf{F} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial \pi_i \, \partial \pi_j} \right\rangle \, , \tag{6.31}$$

where the derivative is to be evaluated at the true parameter value $\boldsymbol{\pi}_0$ (which on average coincides with the maximum-likelihood point if the maximum-likelihood estimator is unbiased, as is the case here). F can be interpreted as an estimate of the inverse covariance matrix of the parameter estimates, provided the likelihood is well approximated by a Gaussian close to the maximum likelihood point. The square root of the determinant of F^{-1} therefore is a good measure of the volume of the 1σ-credible region. For instance, Simon et al. (2004) use this quantity to investigate the gain in cosmological information for different numbers of redshift bins for cosmic shear tomography.

To demonstrate the bias in $\sqrt{\det \mathsf{F}^{-1}}$, we compute the Fisher matrix for the straight line and power law fits using (Tegmark et al. 1997)

$$\mathsf{F}_{ij} = \sum_{\alpha,\beta=1}^{p} \frac{\partial m_\alpha}{\partial \pi_i} \frac{\partial m_\beta}{\partial \pi_j} \, \mathsf{C}_{\alpha\beta}^{-1} \, , \tag{6.32}$$

which is valid if the covariance matrix does not depend on the parameters π_i; $\boldsymbol{m}$ is the model vector.

In Fig. 6.3, we give the ratio of $\sqrt{\det \mathsf{F}^{-1}}$, already computed using the unbiased estimated covariance $\hat{\mathsf{C}}^{-1}$, to the value computed using the true covariance (boxes and dashed lines). One sees that in this case the size of the confidence regions is significantly overestimated, for p/N approaching unity by as much as $\approx 30\%$ for the straight line case, and by a comparable, albeit slightly smaller factor for the power law fit.

6.2.5 Bootstrapping

The derivation of the unbiased estimator $\hat{\mathsf{C}}^{-1}$ rests on the assumptions of Gaussian noise and statistically independent data vectors. To test the performance of this estimator a realistic situation, where one or both of these assumptions may be violated, we make use of an example from the field of weak gravitational lensing. We simulate a weak lensing survey consisting of one single field, containing N_{g} galaxies, which are assigned a random ellipticity ϵ. The two components of the ellipticity are drawn from a Gaussian distribution with dispersion $\sigma_\epsilon/\sqrt{2}$.

The goal of the survey is to measure the shear correlation function $\xi_+(\vartheta)$ (see Sec. 3.3.4) and to fit a model prediction to it. An estimator for ξ_+ is given by (Schneider et al. 2002a)

$$\hat{\xi}_+(\vartheta) = \frac{\sum_{ij} \left(\epsilon_1^{(i)} \epsilon_1^{(j)} + \epsilon_2^{(i)} \epsilon_2^{(j)} \right) \Delta_\vartheta \left(|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j| \right)}{2 n_{\mathrm{p}}(\vartheta)} \, , \tag{6.33}$$

where the galaxies are labelled with i and j and have the angular positions $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$. $\Delta_\vartheta(\phi)$ is unity if $\vartheta - \Delta\vartheta/2 < \phi \le \vartheta + \Delta\vartheta/2$, where $\Delta\vartheta$ is the bin width, and zero otherwise. Finally, $n_{\mathrm{p}}(\vartheta)$ is the number of pairs of galaxies contributing to the correlation function in the bin centred on ϑ (the factor two in the denominator in Eq. 6.33 appears because the sum in the numerator counts every pair of galaxies twice).

We also need the covariance matrix of $\xi_+(\vartheta)$, which, since we only have one measurement, is estimated using the Bootstrapping algorithm (e.g. Efron & Tibshirani 1993): First, we create
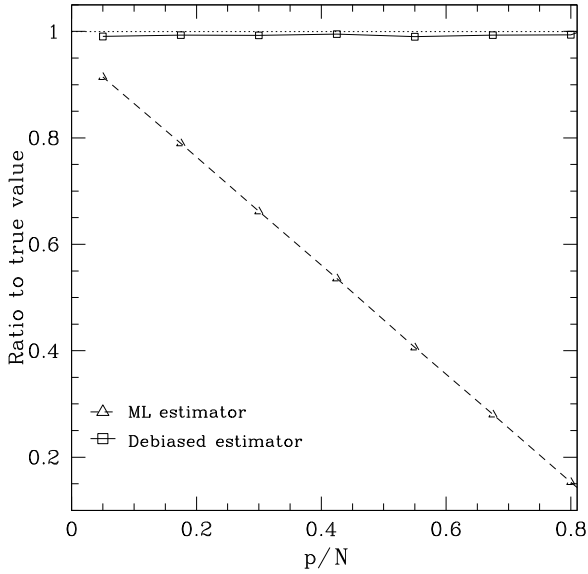
Figure 6.4: *Ratio of the traces of $\hat{\mathsf{C}}_*^{-1}$ and $\hat{\mathsf{C}}^{-1}$ to the trace of Σ^{-1}. The covariances have been estimated using bootstrapping. Error bars are comparable to the symbol size and therefore omitted.*

a catalogue of all $N_\mathrm{p} = N_\mathrm{g}(N_\mathrm{g} - 1)/2$ possible pairs of galaxies in the field. We then create N_bs bootstrap realizations of the survey by repeatedly drawing N_p pairs with replacement from the catalogue. From these, we estimate the mean data vector and the covariance matrix of the shear correlation function. As before, we do this for various numbers of bins, where we record the dependence of the traces of Σ^{-1}, $\hat{\mathsf{C}}_*^{-1}$ and $\hat{\mathsf{C}}^{-1}$ on binning. For the simple case of pure shape noise, which we will consider here, the population covariance is diagonal and can be easily computed using (Schneider et al. 2002a)

$$\Sigma_{ij} = \frac{\sigma_\epsilon^4}{2\,n_\mathrm{p}(\vartheta_i)}\,\delta_{ij}\ ,\tag{6.34}$$

where ϑ_i is the angular position corresponding to the centre of the i-th bin. We precompute the function n_p numerically from a large set of independent data fields for all binning parameters we wish to use in the simulation.

In principle, both of the assumptions made for the derivation of Eq. (6.22) are violated: The noise in the shear correlation function is χ^2-distributed, because $\xi_+ \propto \epsilon\epsilon$, where ϵ is drawn from a Gaussian. However, the number of degrees of freedom of the χ^2-distribution, which equals the number of pairs, is very large, so that it is very well approximated by a Gaussian (central limit theorem). We therefore do not expect any significant influence on the performance of $\hat{\mathsf{C}}^{-1}$. We expect a larger impact by the fact that the data vectors resulting from the bootstrapping procedure are not statistically independent, since different bins necessarily contain the identical galaxy pairs. Strictly speaking, also the requirements for the application of the bootstrap procedure are not met, since the pairs of galaxies which we use to sample the distribution of the shear correlation function are not statistically independent. However, we argue that drawing individual galaxies instead of pairs is not correct, since this would sample the distribution of ϵ, and not the distribution of ξ_+.

The outcome of $\approx 2 \times 10^4$ realizations of this experiment is given in Fig. 6.4, with $N_\mathrm{g} = 500$ and $N_\mathrm{bs} = 40$. The figure shows that, in spite of the correlations among the pairs of galaxies and the data vectors, $\hat{\mathsf{C}}^{-1}$ is wrong by only $\approx 1\%$, and may well be used in bootstrap applications like this.

6.3 The non-Gaussianity of the Cosmic Shear likelihood

In the near future, a new generation of weak lensing surveys like KIDS or Pan-STARRS (Kaiser & Pan-STARRS Collaboration 2005) will allow cosmic shear to be measured with statistical uncertainties much smaller than the systematic errors both on the observational and the theoretical side. Large efforts are now being undertaken to find sources of systematics in the process of shape measurement and shear estimation (e.g. Massey et al. 2007a). In addition, new methods of shape measurement are being explored, such as the shapelet formalism (Refregier & Bacon 2003) or the method proposed in Bernstein & Jarvis (2002).

It is equally important to have accurate theoretical model predictions which can be fit to the expected high-quality measurements. Currently, these models are all based on fitting formulae for the three-dimensional matter power spectrum derived from N-body simulations as given by Peacock & Dodds (1996) and more recently by Smith et al. (2003). However, as we have shown in Sec. 5.2.1, these are only accurate at best to the percent level when compared to ray-tracing simulations based on state-of-the-art N-body simulations such as the Millennium Simulation. Therefore, there is a great need for a large ray-tracing effort to obtain accurate semi-numerical predictions for a range of cosmological parameters.

While a tremendous effort is currently dedicated to the solution of these problems, the actual process of parameter estimation so far has received relatively little attention. Obviously, the statistical data analysis has to achieve the same accuracy as the data acquisition if the aforementioned efforts are not be wasted. The standard procedure to convert measurements of second-order cosmic shear statistics into constraints on cosmological parameters is to write down a likelihood function and to determine the location of its maximum for obtaining estimates of the cosmological parameters of interest. To make this feasible, several approximations are commonly made. Despite the fact that the shear field is non-Gaussian due to non-linear structure growth, lacking an analytical description, the likelihood is most often approximated by a multivariate Gaussian distribution. The covariance matrix for the Gaussian likelihood then remains to be determined, which is an intricate issue by itself. In all previous studies, the dependence of the covariance matrix on cosmological parameters has been ignored. Instead, it was kept fixed to some fiducial cosmological model. While this is not expected to change the degeneracies between different cosmological parameters, it will changes the error estimates perpendicular to the directions of degeneracy (Eifler et al. 2008). There are several approaches to determine the covariance for the fiducial set of parameters: Hoekstra et al. (2006) use the covariance matrix derived for a Gaussian shear field. Although this is rather easy to compute (Joachimi et al. 2008), the errors are strongly underestimated particularly on small scales. Another option is to estimate the covariance from the data itself (e.g. Massey et al. 2007b). This will become sensible and feasible mostly for the upcoming large surveys, which can be safely split into smaller subfields without severely underestimating cosmic variance. A third possibility, which currently seems to be the most accurate, is to measure the covariance matrix from a large sample of ray-tracing simulations. Semboloni et al. (2007) have provided a fitting formula which allows one to transform Gaussian covariances into non-Gaussian ones. Another promising way, which would also easily allow one to take into account the dependence on cosmological parameters, is the semi-analytical computation using the halo model (Cooray & Hu 2001a).

However, all these methods are based on the assumption that the likelihood is well approximated by a Gaussian. In this section, we study the impact of this assumption on the shape of the posterior probability distribution of the matter density parameter Ω_{m} and the power spectrum normalization σ_8. Furthermore, we compute Fisher matrix constraints for the four-dimensional

parameter space spanned by Ω_{m}, σ_8, h_{100} and Ω_Λ. We propose a method to numerically compute the likelihood function from a large set of ray-tracing simulations based on the technique of *Independent Component Analysis* (ICA, e.g. Jutten & Hérault 1991; Comon et al. 1991). ICA is a technique for the separation of independent source signals underlying a set of observed random variables, a statistical method related to factor analysis and Principal Component Analysis (PCA). Since our method depends on certain assumptions, we check our results by comparing to a related, but independent method (Projection Pursuit Density Estimation; Friedman et al. 1984).

6.3.1 The simulations

We have performed ray-tracing trough the the ten CDFS simulations (see Sec. 4.5.2) using `RTsuite` to create 9600 quasi-independent realizations of a field very similar to the Chandra Deep Field South. We have used the redshift distribution

$$p(z_{\mathrm{s}}) = A \left(\frac{z_{\mathrm{s}}}{z_0}\right)^\alpha \exp\left[-\left(\frac{z_{\mathrm{s}}}{z_0}\right)^\beta\right] ,$$

where $z_0 = 1.55$, $\alpha = 0.59$, $\beta = 1.35$ (as estimated from the CDFS source catalogues) and A is a normalization constant. We then created the mock source catalogue by randomly sampling the shear maps from the ray-tracing simulations with $N_{\mathrm{s}} = n_{\mathrm{s}}\Omega^2$ galaxies, where $n_{\mathrm{s}} = 68$ arcmin^{-2} is the number density of sources and $\Omega = 0.5°$ is the side length of the simulated field.

6.3.2 Estimating the likelihood

Let us assume that we have measured the shear correlation functions $\xi_\pm(\theta_i)$ for various angular separation bins θ_i and now wish to infer some parameters $\boldsymbol{\pi}$ of our model $\boldsymbol{m}(\boldsymbol{\pi})$ for $\xi_\pm(\theta_i)$. For what follows, we define the joint data vector $\boldsymbol{\xi} = (\boldsymbol{\xi}_+, \boldsymbol{\xi}_-)^{\mathrm{t}}$, which in total is supposed to have p entries.

Our aim is to compute the posterior likelihood (see Eq. 6.1)

$$p(\boldsymbol{\pi}|\boldsymbol{\xi}) = \frac{p(\boldsymbol{\pi})}{p(\boldsymbol{\xi})} p(\boldsymbol{\xi}|\boldsymbol{\pi}) . \tag{6.35}$$

Hitherto, it has been assumed in the literature that the likelihood $p(\boldsymbol{\xi}|\boldsymbol{\pi})$ is a Gaussian distribution:

$$p(\boldsymbol{\xi}|\boldsymbol{\pi}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\xi} - \boldsymbol{m}(\boldsymbol{\pi}))^{\mathrm{t}} \, \mathsf{C}^{-1}(\boldsymbol{\pi}) \, (\boldsymbol{\xi} - \boldsymbol{m}(\boldsymbol{\pi}))\right] , \tag{6.36}$$

where $\mathsf{C}(\boldsymbol{\pi})$ is the covariance matrix of $\boldsymbol{\xi}$ as predicted by the underlying model. Usually, however, the dependence of the covariance matrix upon cosmological parameters is not taken into account. Rather, the covariance that is computed for a fixed fiducial set of parameters $\boldsymbol{\pi}_0$ is used in Eq. (6.36). Under this approximation, the likelihood is a function of the difference $\boldsymbol{\Delta}(\boldsymbol{\pi}) = \boldsymbol{\xi} - \boldsymbol{m}(\boldsymbol{\pi})$ only:

$$p(\boldsymbol{\xi}|\boldsymbol{\pi}) = L_{\boldsymbol{\pi}_0}[\boldsymbol{\Delta}(\boldsymbol{\pi})] . \tag{6.37}$$

The choice of the functional form of the likelihood as given by Eq. (6.36) is only approximate. Since the underlying shear field in the correlation function measurement becomes non-Gaussian

in particular on small scales due to non-linear structure formation, there is no good reason to expect the distribution of the shear correlation function to be Gaussian. Our aim therefore is to use our large sample of ray-tracing simulations to estimate the likelihood $\mathcal{L}$ and explore the effects of the deviations from a Gaussian shape on cosmological parameter constraints. However, we have to sustain the approximation that the functional form of the likelihood does not depend on cosmology in order to keep computation time manageable.

Our ray-tracing simulations were all done for identical cosmological parameters, which comprise our fiducial parameter vector $\boldsymbol{\pi}_0$. Thus, as in Eq. (6.37) the likelihood depends on cosmology only through the difference $\boldsymbol{\Delta}(\boldsymbol{\pi}) = \boldsymbol{\xi} - \boldsymbol{m}(\boldsymbol{\pi})$. Since L is the probability of obtaining the data $\boldsymbol{\xi}$ given the model $\boldsymbol{m}$, we in principle have to estimate the p-dimensional distribution of $\boldsymbol{\xi}$ from our sample of N simulations. However, due to the high dimensionality of the problem, a brute force approach to estimate the full joint distribution is hopeless. The problem would simplify considerably if we could find a transformation

$$\boldsymbol{s} = \boldsymbol{f}\left[\boldsymbol{\Delta}(\boldsymbol{\pi})\right] \;\;, \tag{6.38}$$

such that

$$p_s(\boldsymbol{s}|\boldsymbol{\pi}_0) = \prod_{i=1}^{n_{\mathrm{IC}}} p_{s_i}(s_i|\boldsymbol{\pi}_0) \;\;. \tag{6.39}$$

Here, $\boldsymbol{f}$ is in general a mapping from $\mathbb{R}^p$ to $\mathbb{R}^{n_{\mathrm{IC}}}$ ($n_{\mathrm{IC}} \leq p$) and $\boldsymbol{s} \in \mathbb{R}^{n_{\mathrm{IC}}}$ is our new data vector. This would reduce the problem to estimating n_{IC} one-dimensional probability distributions instead of a single p-dimensional one. Eq. (6.39) is equivalent to the statement that we are looking for a new set of basis vectors of $\mathbb{R}^{n_{\mathrm{IC}}}$ in which the components s_i of the shear correlation function are statistically independent. It is virtually impossible to find the (in general nonlinear) mapping $\boldsymbol{f}$. However, it is possible to make progress if we make the ansatz that $\boldsymbol{f}$ is linear:

$$\boldsymbol{s} = \mathsf{A}\boldsymbol{\Delta}(\boldsymbol{\pi}) \;\;, \tag{6.40}$$

where $\mathsf{A} \in \mathbb{R}^{n_{\mathrm{IC}} \times p}$ is the transformation or "un-mixing" matrix.

Our likelihood estimation procedure is as follows: the first step is to remove first-order correlations from the data vector by performing a PCA (e.g. Press et al. 1992). This yields a basis in which the components of $\boldsymbol{\xi}$ are uncorrelated. If we knew that the distribution of $\boldsymbol{\xi}$ were Gaussian, this would be sufficient, because in this case uncorrelatedness is equivalent to statistical independence. However, for a general distribution, uncorrelatedness is only a necessary condition for independence. Since we suspect that the likelihood is non-Gaussian, a second change of basis, determined by the ICA technique (described in detail in the next section), is carried out which then results in the desired independence. We use a kernel density method (see e.g. Hastie et al. 2001; Venables & Ripley 2002, and references therein) to estimate and tabulate the one-dimensional distributions $p_{s_i}(s_i|\boldsymbol{\pi}_0)$ in this new basis. The density estimate is constructed by smoothing the empirical distribution function of the observations of s_i,

$$p_{s_i}^{\mathrm{emp}}(x) = \sum_{j=1}^{N} \delta_{\mathrm{D}}\left(x - s_i^{(j)}\right) \;\;, \tag{6.41}$$

where $s_i^{(j)}$ is the j-th of N observations of s_i, with a smooth kernel K. The estimate $\hat{p}_{s_i}$ of the desired density p_{s_i} then is given by

$$\hat{p}_{s_i}(x) = \frac{1}{Nb} \sum_{j=1}^{N} K\left(\frac{x - s_i^{(j)}}{b}\right) \;\;, \tag{6.42}$$

where b is the bandwidth. For the kernel K we use a Gaussian distribution. It has been shown that the shape of the kernel K is of secondary importance for the quality of the density estimate; much more important is the choice of the bandwidth b. If b is too small, $\hat{p}_{s_i}$ is essentially unbiased, but tends to have a high variance because the noise is not properly smoothed out. On the other hand, choosing a bandwidth that is too large results in a smooth estimate with low variance, but a higher bias, because real small scale features of the probability density are smeared out. Our choice of the bandwidth is based on the "rule of thumb" (Scott 1992).

Constraints on cosmological parameters can then be derived as follows: we transform our set of model vectors and the measured correlation function to the new ICA basis:

$$\breve{m}(\pi) = \mathsf{A}\,m(\pi)\ ,\tag{6.43}$$

$$\breve{\xi} = \mathsf{A}\,\xi\ ,\tag{6.44}$$

so that $s = \mathsf{A}\,[\xi - m(\pi)]$. The ICA posterior distribution is then given by

$$p(\pi|\xi) \propto p(\pi) \prod_{i=1}^{n_{\mathrm{IC}}} p_{s_i}(\breve{\xi}_i - \breve{m}_i(\pi)|\pi_0)\ .\tag{6.45}$$

6.3.3 Independent Component Analysis

Information theory

Before discussing the ICA algorithm in detail, we briefly review some relevant aspects of information theory. A more thorough discussion of these concepts can be found in, e.g., MacKay (2003) or Hyvärinen et al. (2001).

The *entropy* of a discrete-valued random variable X is defined as

$$H_X = -\sum_i P(X = x_i)\, \log P(X = x_i)\tag{6.46}$$

where the x_i are the possible values of X and $P(X = x_i)$ is the probability of the outcome $X = x_i$. The entropy can be interpreted as the information content of X: the more unpredictable the value of X is, the larger is the entropy. In fact, it can be shown that the entropy assumes its maximum if $P(X = x_i) = P(X = x_j)$ for all i, j. On the other hand, the entropy is very small if $P(X = x_i)$ is close to zero for all but one i. Therefore, a low value of H means that there is little randomness in X. Eq. (6.46) can be extended to a continuous random vector y by defining the differential entropy

$$H_y = -\int \mathrm{d}y'\, p(y')\, \log p(y')\ .\tag{6.47}$$

Here, p is the probability distribution function of y. Unlike Eq. (6.46), the differential entropy is no longer positive semi-definite. If we introduce a new random vector $z = \mathsf{M}y$, where M is an invertible matrix, it can be seen immediately that the differential entropy of z is given by

$$H_z = H_y + \log|\det \mathsf{M}|\ .\tag{6.48}$$

A "distance measure" between two probability distribution functions p and q is provided by the *Kullback-Leibler divergence* (Kullback & Leibler 1951):

$$D_{\mathrm{KL}}[p, q] = \int \mathrm{d}x\, p(x)\, \log \frac{p(x)}{q(x)}\ .\tag{6.49}$$

The interpretation of the K-L divergence D_{KL} as a distance is sensible because it is non-negative and zero only if $p \equiv q$, albeit not symmetric.

The *mutual information* between the n components of a random vector $\boldsymbol{x}$ is defined to be

$$I_{x_1,\ldots,x_n} = \sum_{i=1}^{n} H_{x_i} - H_{\boldsymbol{x}} , \tag{6.50}$$

where H_{x_i} is the entropy of the continuous random variable x_i (the i-th component of $\boldsymbol{x}$) only, whereas $H_{\boldsymbol{x}}$ is the entropy of the full random vector $\boldsymbol{x}$.

Mutual information can be expressed in terms of the K-L divergence: if the components of $\boldsymbol{x}$ were statistically independent, their joint density would be given by $p'(\boldsymbol{x}) = p_1(x_1) \cdots p_n(x_n)$, where $p_i(x_i)$ is the (marginalized) distribution function of x_i alone. It is therefore obvious that a measure of the independence of the x_i can be constructed by computing the K-L divergence of the factorized density p' and the full joint density p:

$$
\begin{aligned}
D_{\text{KL}}[p, p'] &= \int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{p_1(x_1) \cdots p_n(x_n)} \\
&= \int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \log p(\boldsymbol{x}) - \sum_{i=1}^{n} \int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \log p_i(x_i) \\
&= -H_{\boldsymbol{x}} - \sum_{i=1}^{n} \int \mathrm{d}x_i\, p_i(x_i) \log p_i(x_i) \\
&= I_{x_1,\ldots,x_n} ,
\end{aligned}
\tag{6.51}
$$

where in the second-last step we have used that

$$
\begin{aligned}
\int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \log p_i(x_i) &= \int \mathrm{d}x_i\, \log p_i(x_i) \int \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\mathrm{d}x_{i+1} \cdots \mathrm{d}x_n\, p(\boldsymbol{x}) \\
&= \int \mathrm{d}x_i\, \log p_i(x_i)\, p_i(x_i) = -H_{x_i} .
\end{aligned}
\tag{6.52}
$$

From the positive semi-definiteness of D_{KL} it therefore follows that $I(x_1, \ldots, x_n) \geq 0$, where the equality only holds if the x_i are statistically independent.

In what follows, we shall need a measure of non-Gaussianity of a probability distribution function. It can be shown that the Gaussian distribution has the largest entropy among all distributions with fixed variance (in other words, the Gaussian distribution is the maximum-entropy distribution for this case). Therefore, one defines the *negentropy*:

$$J_{\boldsymbol{x}} = H_{\boldsymbol{x}_{\text{Gauss}}} - H_{\boldsymbol{x}} , \tag{6.53}$$

where $\boldsymbol{x}_{\text{Gauss}}$ is a Gaussian random vector with the same covariance matrix as $\boldsymbol{x}$. The negentropy is always non-negative and zero if $\boldsymbol{x}$ is Gaussian.

The method

We now describe the ICA method (Hyvärinen et al. 2001; Hyvärinen & Oja 2000), which we use to find the new basis in $\mathbb{R}^{n_{\text{IC}}}$ in which the components of $\boldsymbol{\Delta}$ are (approximately) statistically independent.

ICA is best introduced by assuming that the data at hand were generated by the following linear model (using the notation of Eq. 6.40):

$$\boldsymbol{\Delta} = \mathsf{M}\boldsymbol{s} \ , \tag{6.54}$$

where $\boldsymbol{s}$ is a vector of statistically independent source signals with non-Gaussian probability distributions and M is the $p \times n_{\mathrm{IC}}$ mixing matrix. The goal of ICA is then to estimate both M and $\boldsymbol{s}$ from the data.

For simplicity, we will from now on only consider the case $n_{\mathrm{IC}} = p$. In this case, the mixing matrix M is simply the inverse of the un-mixing matrix A in Eq. (6.40).

An intuitive, though slightly hand-waving way to understand how this estimation problem can be solved is to note that a set of linear combinations y_i of independent, non-Gaussian random variables x_j will usually have distributions that are more Gaussian than the original distributions of the x_j (by virtue of the central limit theorem). Reversing this argument, this suggests that the x_j could be recovered from a sample of the y_i by looking for linear combinations of the y_i that have the least Gaussian distributions. These linear combinations will also be close to statistically independent.

For a more rigorous justification, consider the random vector of source signals $\boldsymbol{s} = \mathsf{M}^{-1}\boldsymbol{\Delta}$. Using the transformation property of the entropy as given in Eq. (6.48), we can write the mutual information of the s_i as

$$I_{s_1,\ldots,s_{n_{\mathrm{IC}}}} = \sum_{i=1}^{n_{\mathrm{IC}}} H_{s_i} - H_{\boldsymbol{\Delta}} - \log|\det\mathsf{M}|^{-1} \tag{6.55}$$

As will be discussed below, the each source signal s_i can only be determined up to a multiplicative constant using ICA. We choose these factors such that $\langle\boldsymbol{ss}^{\mathrm{t}}\rangle = \mathsf{E}$, where E is the unit matrix. In this case, $\det\mathsf{M}$ is constrained to be constant, because $\det\langle\boldsymbol{ss}^{\mathrm{t}}\rangle = \det\mathsf{E} = \det\mathsf{M}^{-1}\langle\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathrm{t}}\rangle\det\mathsf{M}^{-1}$, and $\langle\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathrm{t}}\rangle$ does not depend on M. Thus, the only term in Eq. (6.55) that depends on the choice of the transformation matrix is the first. On the other hand, note from Eq. (6.53) that for a random vector of unit variance, negentropy and entropy only differ by a constant (the entropy of the corresponding Gaussian random variable) and the sign. We can therefore write Eq. (6.55) as

$$I_{s_1,\ldots,s_p} = \mathrm{const.} - \sum_{i=1}^{p} J_{s_i} \ . \tag{6.56}$$

This shows that finding a linear transformation M that minimizes the mutual information (as a measure of statistical dependence) of the resulting source vector $\boldsymbol{s}$ is equivalent to maximizing the negentropy of the components of $\boldsymbol{s}$ (a measure of non-Gaussianity). This agrees with the intuitive argument in the previous paragraph.

The ICA algorithm consists of two parts, the first of which is a preprocessing step: after subtracting the mean $\bar{\boldsymbol{\Delta}} = \langle\boldsymbol{\Delta}\rangle$ from $\boldsymbol{\Delta}$, the data is *whitened*, i.e. a linear transformation $\tilde{\boldsymbol{\Delta}} = \mathsf{L}\boldsymbol{\Delta}$ is introduced such that $\langle\tilde{\boldsymbol{\Delta}}\tilde{\boldsymbol{\Delta}}^{\mathrm{t}}\rangle = \mathsf{E}$. This can be achieved by computing the eigendecomposition of the covariance matrix $\mathsf{C} = \mathsf{U}\mathsf{D}\mathsf{U}^{\mathrm{t}}$ of $\boldsymbol{\Delta}$ and by choosing $\mathsf{L} = \mathsf{D}^{-1/2}\mathsf{U}^{\mathrm{t}}$, where U is orthonormal, $\mathsf{D} = \mathrm{diag}(d_1,\ldots,d_p)$ and $d_i \geq 0$ for all i. The effect of the whitening is that the new mixing matrix $\tilde{\mathsf{M}} = \mathsf{L}\mathsf{M}$ between $\tilde{\boldsymbol{\Delta}}$ and $\boldsymbol{s}$ is orthogonal. This can be seen as follows: $\mathsf{E} = \langle\tilde{\boldsymbol{\Delta}}\tilde{\boldsymbol{\Delta}}^{\mathrm{t}}\rangle = \tilde{\mathsf{M}}\langle\boldsymbol{ss}^{\mathrm{t}}\rangle\tilde{\mathsf{M}}^{\mathrm{t}}$, and since we have $\langle\boldsymbol{ss}^{\mathrm{t}}\rangle = \mathsf{E}$, the claim follows.

After the preprocessing, the components of $\tilde{\boldsymbol{\Delta}}$ are uncorrelated. This would be equivalent to statistical independence if their distributions were Gaussian. However, as this is not the case here, a further step is needed. It consists of finding a new set of orthogonal basis vectors $\boldsymbol{w}_i$ (the row vectors of $\tilde{\mathsf{M}}$) such that the distributions $p_{z_i}(z_i)$ of

$$z_i = \tilde{\boldsymbol{\Delta}}_i \cdot \boldsymbol{w}_i \tag{6.57}$$

maximize the negentropy of the z_i. Starting from randomly chosen initial directions $\boldsymbol{w}_i$, the algorithm tries to achieve this goal iteratively (in practice, it is sufficient to use a simple approximation to the negentropy). For more details on the practical implementation, the reader is referred to Hyvärinen et al. (2001).

ICA suffers from several ambiguities, none of which, however, is crucial for this work. First of all, the variances (weights) of the source signals cannot be determined, since any prefactor λ to the signal s_i can be cancelled by multiplication of the corresponding column of the mixing matrix by $1/\lambda$. Secondly, the order of the independent components is not determined, since any permutation of the s_i can be accommodated by corresponding changes to M. Thirdly, ICA does not yield a unique answer if at least some of the s_i are Gaussian – the subset of Gaussian signals is only determined up to an orthogonal transformation. This is not an issue in our context, since the Gaussian signals will be uncorrelated thanks to the preprocessing steps, and uncorrelatedness implies statistical independence for Gaussian random variables.

Several interpretations of ICA and algorithms exist and are described in detail in Hyvärinen et al. (2001). In this work, we use an implementation of the `fastICA` algorithm (Hyvärinen & Oja 1997) for the `R` language (R Development Core Team 2007)[1].

6.3.4 Tests

Statistical biases

In this section, we present the results of a number of tests we have performed to insure that the results obtained with the ICA method are not affected by convergence issues or statistical biases of any kind.

As discussed in Sec. 6.3.3, the `fastICA` algorithm requires a set of randomly chosen directions $\boldsymbol{w}_i$ as initial conditions. It then iteratively computes corrections to these vectors in order to increase the negentropy of the projections of the data vectors onto these directions (Eq. 6.57), followed by an orthonormalization step. Therefore, it is not clear a priori whether the algorithm will settle in the same negentropy maxima for different sets of initial vectors. This concern is backed by the fact that at least some of the $p_{s_i}(s_i|\boldsymbol{\pi}_0)$ might be very close to Gaussian, which might hamper convergence even further due to the inability of ICA to uniquely distinguish between Gaussian source signals. We have therefore tested whether we obtain the same set of basis vectors from a large number of different initial vectors $\boldsymbol{w}_i$. Indeed, we reliably recover those basis vectors for which the distribution of $p_{z_i}(z_i = \tilde{\boldsymbol{\Delta}} \cdot \boldsymbol{w}_i)$ departs significantly from a Gaussian. As expected, the directions leading to a rather Gaussian p_{z_i} are different for different starting values. As discussed before, this is no reason for concern, since projections z_j of the data onto these "Gaussian" directions have been decorrelated already through the whitening transformation. Since the ICA mixing matrix $\tilde{\mathsf{M}}$ is orthogonal, the observed mixing of these components will preserve the uncorrelatedness. We find indeed that the posterior distributions derived using our algorithm do not differ notably when using different starting vectors. This
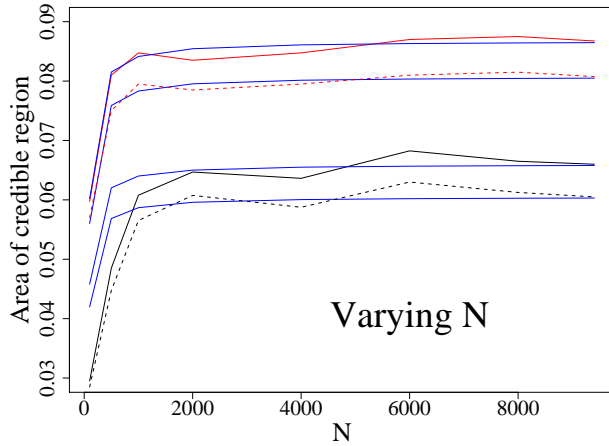
---

[1] `http://www.r-project.org/`

Figure 6.5: *Area of the 68% (dashed lines) and 99% (solid lines) credible regions in the Ω_{m}-σ_8-plane as function of the sample size N, for the Gaussian likelihood (red) and the likelihood computed using the ICA algorithm (black). Blue lines are the predicted areas based on Eq. (6.58).*

is even true if the `fastICA` algorithm does not formally converge (i.e. when the differences of some of the basis vectors between two iterations is not small): after a few hundred iterations, the non-Gaussian directions are determined and do not change anymore. The reason for not reaching convergence is that the algorithm still tries to find negentropy maxima in the subspace of Gaussian directions.

As has been noted in Sec. 6.2.1, statistical biases can become significant already for the Gaussian approximation of the likelihood (Eq. 6.36): care has to be taken if the covariance matrix of the correlation function (given on p bins) is estimated from a finite set of N simulations or observations. Inverting the estimated covariance yields a biased estimate of the inverse (see Eq. 6.22):

$$\left\langle \hat{\mathsf{C}}^{-1} \right\rangle = \frac{N-1}{N-p-2}\, \Sigma^{-1} \text{ for } p < N-1\,, \tag{6.58}$$

where $\hat{\mathsf{C}}$ is the estimated and Σ the true covariance matrix. This bias leads to an underestimation of the size of credible regions by a factor of $(N-p-2)/(N-1) \approx 1 - p/N$. We suspect that a similar bias occurs in our likelihood estimation procedure. In Fig. 6.5, we therefore plot the area of the 68% and 99% credible regions of the posterior distribution for Ω_{m} and σ_8 (keeping all other cosmological parameters fixed to their fiducial values) as functions of the number N of observations of the correlation functions used to estimate the ICA transformation (black curves). To exclude noise effects from the analysis, we use the theoretical prediction of the correlation function for the fiducial cosmological parameters as data vector. We set $p = 30$ throughout. For comparison, we also show the areas computed using the Gaussian likelihood (red curves). In the latter case, the bias predicted by Eq. (6.58) is clearly visible as a decrease of the area when N becomes small. The ICA method suffers from a similar bias, although the behaviour at small N seems to be slightly different. More important, though, is the fact that this bias is unimportant for reasonably large sample sizes ($N \gtrsim 2000$). Since we always use the full sample ($N = 9600$) in the following, this bias is completely negligible.
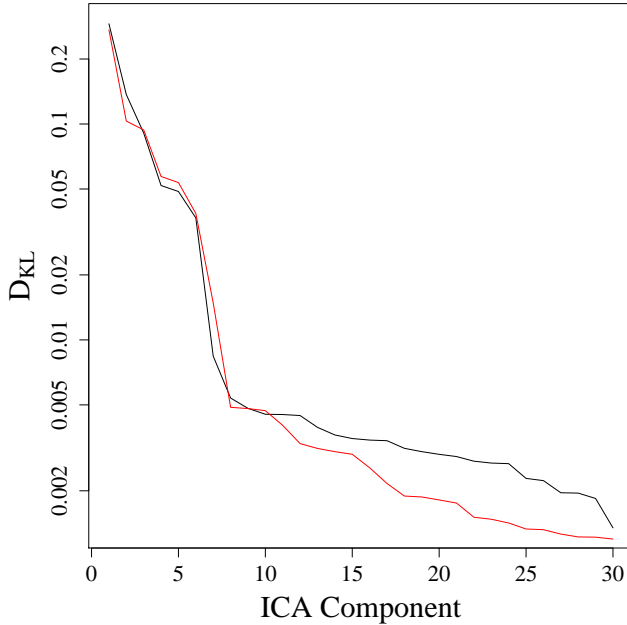
Figure 6.6: *Kullback-Leibler divergences between the estimated probability distributions of the ICA components and the best fitting Gaussian distributions. For the red curve, Eq. (3.26) was used to compute the observed galaxy ellipticities, and for the black curve, the weak lensing limit of this equation was used. The components have been ordered according to their K-L divergence.*

Non-Gaussianity from the reduced shear

The intrinsic ellipticity $\epsilon^{(\mathrm{s})}$, reduced shear $g = \gamma/(1-\kappa)$ and observed image ellipticity ϵ are related by (see Eq. 3.26):

$$\epsilon = \begin{cases} \frac{\epsilon_{\mathrm{s}}+g}{1+g^*\epsilon_{\mathrm{s}}} & \text{if } |g| \leq 1 \\ \frac{1+g\epsilon_{\mathrm{s}}^*}{\epsilon_{\mathrm{s}}^*+g^*} & \text{if } |g| > 1 \ . \end{cases} \tag{6.59}$$

Since this relation is non-linear, it can introduce additional non-Gaussianity to the likelihood, even if the distribution of the intrinsic ellipticities is Gaussian. To test this, we have created two mock galaxy catalogues from our ray-tracing simulations, one using Eq. (3.26) to compute the observed galaxy ellipticities (case A), the other one using the linear weak lensing limit of this equation, given by $\epsilon = \gamma + \epsilon^{(\mathrm{s})}$ (case B). We use both catalogues to estimate the ICA likelihood and compare the resulting p_{s_i}. A convenient measure of non-Gaussianity is the Kullback-Leiber divergence between the kernel-density estimate of p_{s_i} (Eq. 6.42) and the best-fitting Gaussian distribution. In Fig. 6.6, we plot the K-L divergences for the ICA components for both cases. We do not find any significant additional non-Gaussianity that is caused by Eq. (3.26), and also do not see any noteworthy differences in the parameter constraints based on the likelihoods estimated from the two galaxy catalogues.

ICA performance

As a first check of how well the ICA does in separating the independent components of the correlation function, we compare the probability distributions of the twelve most non-Gaussian components with respect to the PCA and ICA basis vectors (Fig. 6.7). As expected, using the ICA we find many more significantly non-Gaussian PDFs than just performing a PCA.
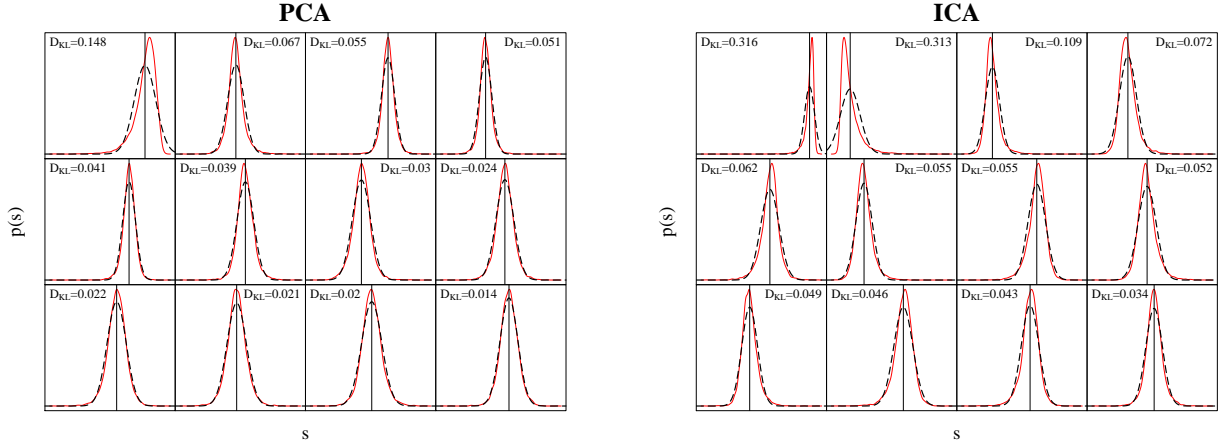
Figure 6.7: *Distributions of the components of the shear correlation function with respect to the PCA basis (left panel) and the ICA basis (right panel). Since the scales of the s_i are irrelevant, we omit the axis labels except for the location of $s_i = 0$, which is indicated by vertical black lines. The solid red lines are the kernel density estimates, the dashed black lines are the best-fitting Gaussian distributions. For each component, we also give the Kullback-Leibler divergence (Eq. 6.49) between the Gaussian and the measured distributions as a measure of non-Gaussianity.*
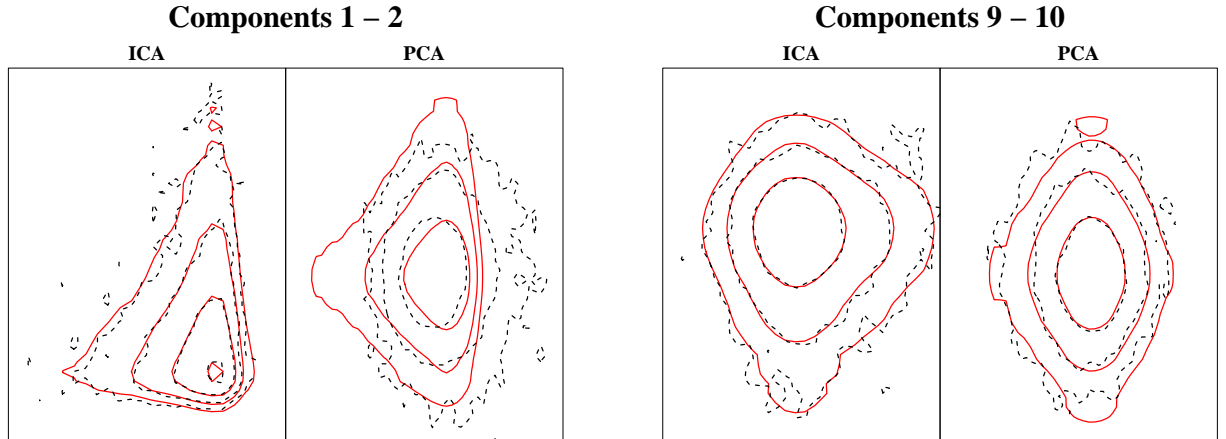
Figure 6.8: *Comparison of the joint distributions $p(s_i, s_j)$ (black dashed contours) and the product $p_{s_i}(s_i) \, p_{s_j}(s_j)$ (solid red contours) for the two most non-Gaussian components ($(i, j) = (1, 2)$) and two rather Gaussian ones ($(i, j) = (9, 10)$). The labels i and j correspond to the ordering of components of Fig. 6.7. In the right panel of each plot, the distributions with respect to the PCA basis vectors are shown and in the left panel, the distributions in the ICA basis are displayed. Statistical independence is indicated by $p(s_i, s_j) = p_{s_i}(s_i) \, p_{s_j}(s_j)$.*

Furthermore, the tendency of the PDFs to be more peaked than a Gaussian, which is visible already in the PCA basis, becomes much clearer. This trend persists also for the remaining components not shown in the figure: while several PCA components are indistinguishable from a Gaussian, no such PDF is found in the ICA case. This shows that the ICA indeed yields an improved separation of non-Gaussian components of the shear correlation function.

Our method to estimate the likelihood crucially depends on the assumption that a linear transformation makes the components of the shear correlation vectors statistically independent. A necessary condition for mutual statistical independence of all s_i is to compare the joint pairwise distributions $p(s_i, s_j)$ to the product distributions $p_{s_i}(s_i) p_{s_j}(s_j)$. The components i and j are pairwise statistical independent if $p(s_i, s_j) = p_{s_i}(s_i) p_{s_j}(s_j)$. We estimate $p(s_i, s_j)$ using a two-dimensional extension (using a bi-variate Gaussian kernel) of the kernel density method given by Eq. (6.42). We give two examples in Fig. 6.8, where we compare the joint and product distributions of the two most-non-Gaussian components and two nearly Gaussian components. As expected, a simple PCA is not enough to achieve pairwise statistical independence in the non-Gaussian case. Only after performing the ICA, pairwise independence is achieved. In addition, Fig. 6.8 nicely illustrates the futility of an attempt to estimate the full p-dimensional likelihood directly: the two-dimensional estimates of the joint densities are already significantly noisier than the product densities.

A more rigorous test for mutual statistical independence for the multivariate, continuous case was proposed by Chiu et al. (2003). It is based on the observation that if x is a continuous random variable and $P(x)$ is its cumulative distribution function (CDF), then $z = P(x)$ is uniformly distributed in $[0, 1]$. If we are given a set of statistically independent random variables s_i, this means that the joint distribution of $z_i = P_i(s_i)$, where again P_i is the CDF of s_i, is uniform in the multidimensional unit cube. On the other hand, if the assumption of statistical independence of the s_i is violated, the joint density $p_{\boldsymbol{z}}$ of the z_i is given by

$$
\begin{aligned}
p_{\boldsymbol{z}}(\boldsymbol{z}) &= p_{\boldsymbol{z}}\left[P_1(s_1), \ldots, P_n(s_n)\right] \\
&= p_{\boldsymbol{s}}(s_1, \ldots, s_n) \left|\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{s}}\right|^{-1} \\
&= \frac{p_{\boldsymbol{s}}(s_1, \ldots, s_n)}{\prod_{i=1}^{n} p_i(s_i)} .
\end{aligned}
\tag{6.60}
$$

Here, $p_i(s_i)$ is the distribution function of s_i only and $p_{\boldsymbol{s}}$ is the joint distribution function of $s_1, \ldots, s_n$. This means that the joint distribution of the z_i is not uniform if the s_i are statistically dependent. Therefore, we can test if the s_i we obtain from the ICA procedure are indeed independent by computing their empirical cumulative distribution functions, carrying out the above transformation and finally testing for multivariate uniformity. Such a test was described in Liang et al. (2001), a discussion of which, however, is considerably beyond the scope of this thesis. Applying the test to the s_i that we have obtained from our ICA procedure, we have to reject statistical independence at 99% confidence. This means that the ICA does not remove all dependencies between the components of the shear correlation function. This result, however, does not give an indication of how these residual dependencies affect our likelihood estimate and the conclusions regarding constraints on cosmological parameters.

We therefore compare the constraints derived from the ICA likelihood with the constraints from the likelihood estimated using an alternative method, called *projection pursuit density estimation* (Friedman et al. 1984), which we describe in detail in App. C. This method is free from any assumptions regarding statistical independence and therefore provides an an ideal
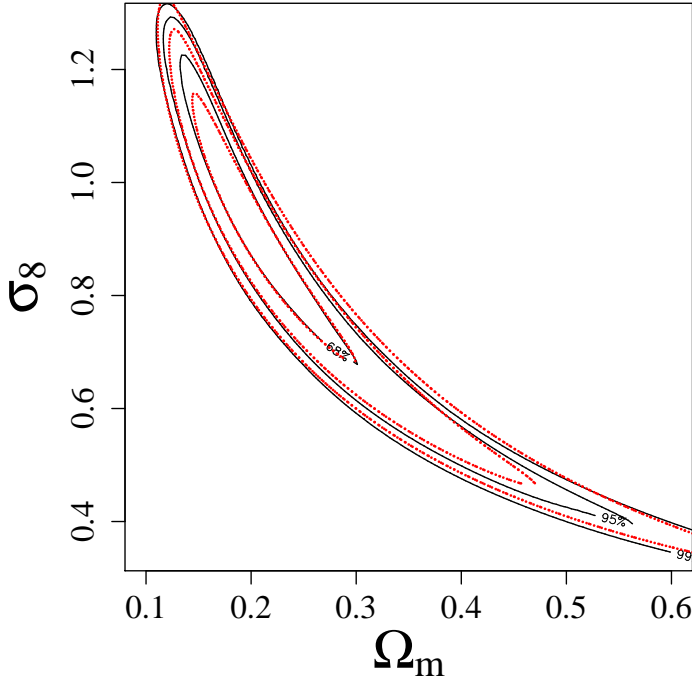
Figure 6.9: *Comparison of the posterior likelihoods for $(\Omega_{\mathrm{m}}, \sigma_8)$, computed using the ICA likelihood (black contours) and the PPDE likelihood (red contours). Shown are the contours of the 68%, 95% and 99% credible regions.*

cross-check for the ICA method. For the comparison, we have computed the shear correlation functions with $p = 10$, and we also use $n_{\mathrm{IC}} = 10$ independent components. The resulting contours in the Ω_{m}-σ_8-plane are shown in Fig. 6.9. Both posterior likelihoods are very similar, although the credible regions of the PPDE posterior have a slightly smaller area than the contours of the ICA posterior (which actually supports the findings presented in the next section). Given the good agreement of the two methods, we will henceforth only make use of the ICA procedure, which is considerably faster and numerically less contrived than PPDE.

6.3.5 Results on the posterior

The most interesting question is how much the posterior distribution computed from the non-Gaussian ICA likelihood will differ from the Gaussian approximation. We have investigated this for the case of the CDFS and the parameter set $(\Omega_{\mathrm{m}}, \sigma_8)$. For the data vector, we do not use the correlation functions from our simulations, but take the theoretical prediction for our fiducial parameter set instead. This allows us to study the shape of the posterior likelihood independent of noise in the data and biases due to the fact that the fiducial theoretical model does not quite match the mean correlation function from the simulations. In Fig. 6.10, we show the contours of the posterior computed in this way from the shape-noise-free likelihood (left panel) and the likelihood estimated from simulations with $\sigma_{\epsilon} = 0.45$ (right panel). The shape of the ICA posterior is different from that of the Gaussian approximation in three respects: it is steeper, the maximum is shifted towards higher σ_8 and lower Ω_{m}, and the contours are slightly tilted. The first two differences can be traced back to the shape of the distributions of the individual ICA components (Fig. 6.7): most the distribution functions are generally slightly steeper than a Gaussian and most of the non-Gaussian components are in addition strongly skewed, thus shifting the peak of the posterior. Generally, these differences are more pronounced in the direction of the Ω_{m}–σ_8-degeneracy and towards lower values of both parameters, where the
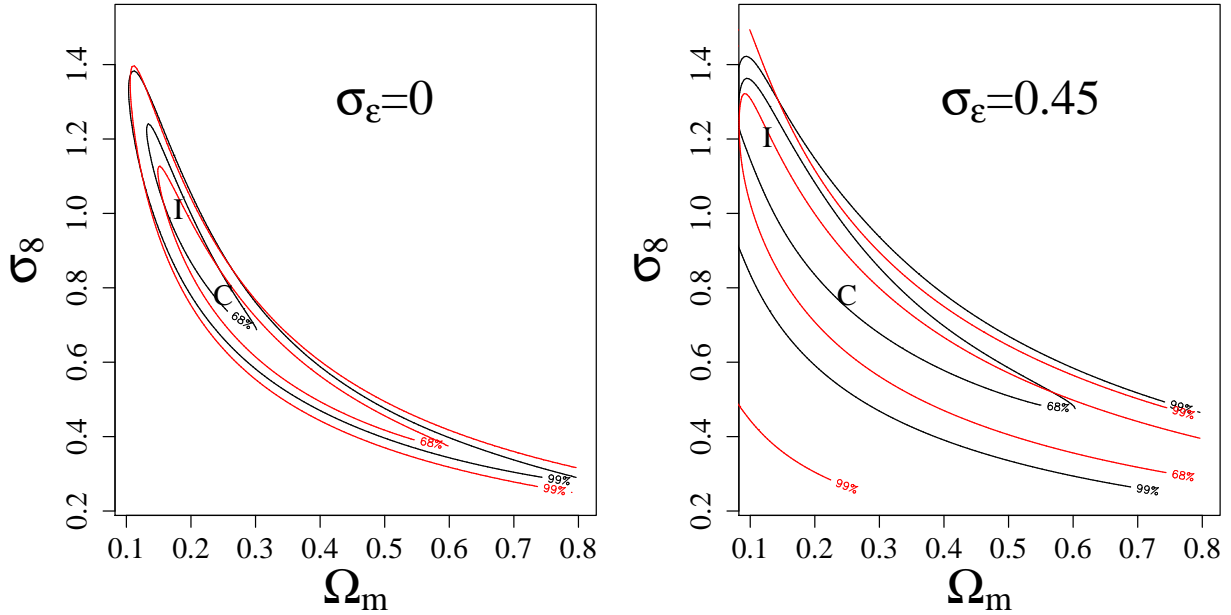
Figure 6.10: *Comparison of the posterior likelihoods for $(\Omega_{\mathrm{m}}, \sigma_8)$, computed using the ICA likelihood (thick black contours) and the Gaussian approximation (dotted red contours). Shown are the contours of the 68% and 99% credible regions. The mode of the ICA posterior is denoted by I. The mode of the posterior based on the Gaussian likelihood coincides with the fiducial parameter set and is marked by the symbol C.*

posterior is shallower. Note that the differences between the Gaussian and the ICA likelihood are much larger than the differences between the ICA and the PPDE likelihood shown in Fig. 6.9.

Of more practical relevance is how the parameter constraints change when the ICA likelihood is used for the analysis of large weak lensing surveys. Here, we consider surveys consisting of N_{f} CDFS-like fields. Bayesian theory states that if N_{f} is large enough, the posterior probability distribution of the parameters becomes Gaussian, centred on the true parameter values, with covariance matrix $(N_{\mathrm{f}} \mathsf{F})^{-1}$ (e.g. Gelman et al. 2004). Here, F is the Fisher matrix (see Eq. 6.31), which can be written as (Kendall et al. 1987)

$$\mathsf{F}_{\alpha\beta} = \left\langle \frac{\partial \log L}{\partial \pi_\alpha} \frac{\partial \log L}{\partial \pi_\beta} \right\rangle \ , \tag{6.61}$$

where $\langle \cdot \rangle$ denotes the expectation value with respect to the likelihood function. If the likelihood is Gaussian and if the covariance matrix C does not depend on cosmology, one can show that (Tegmark et al. 1997)

$$\mathsf{F}_{\alpha\beta} = \sum_{i,j} \mathsf{C}_{ij}^{-1} \frac{\partial m_i(\boldsymbol{\pi})}{\partial \pi_\alpha} \frac{\partial m_j(\boldsymbol{\pi})}{\partial \pi_\alpha} \ . \tag{6.62}$$

Eq. (6.61) provides us with a way to estimate the Fisher matrix for the non-Gaussian likelihood. For each ray-tracing realization of the CDFS, we compute the logarithm of the posterior distribution $\log p(\boldsymbol{\pi}|\boldsymbol{\xi})$ and its derivatives (by finite differencing; see e.g. Abramowitz & Stegun 1964) with respect to the cosmological parameters at the fiducial parameter values. Since we use uniform priors for all cosmological parameters, the derivatives of the log-posterior are identical
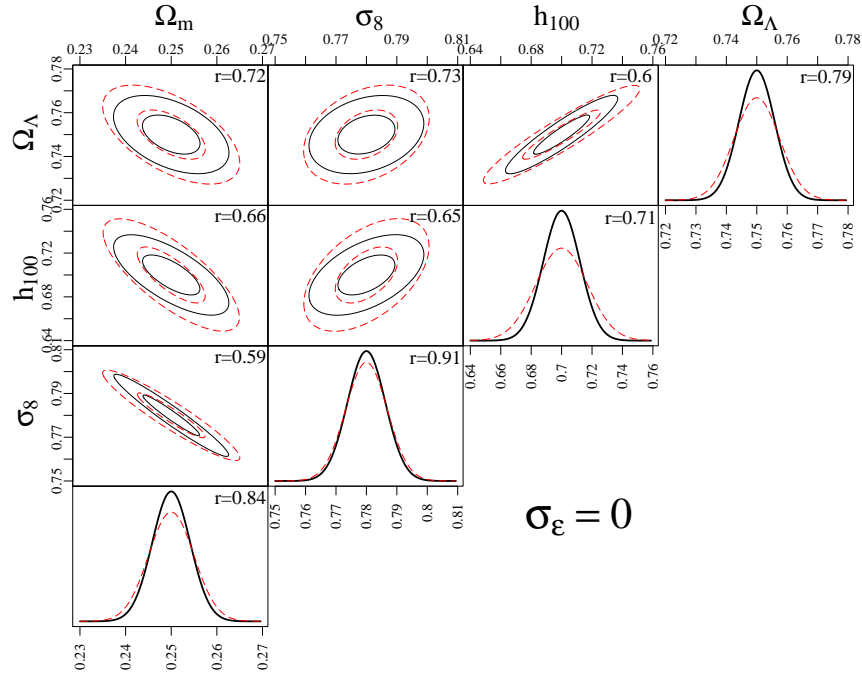
Figure 6.11: *Fisher matrix constraints for a hypothetical* 1500-deg^2 *survey without shape noise. The plots on the diagonal show the 1D marginals, the off-diagonal plots the 2D marginals derived from the full 4D posterior. The red dashed (black solid) lines/contours have been computed using the Fisher matrix of the Gaussian likelihood (the ICA likelihood). In each plot, we give the ratio r of the area of the 68% credible intervals derived from the ICA likelihood to those derived from the Gaussian likelihood.*
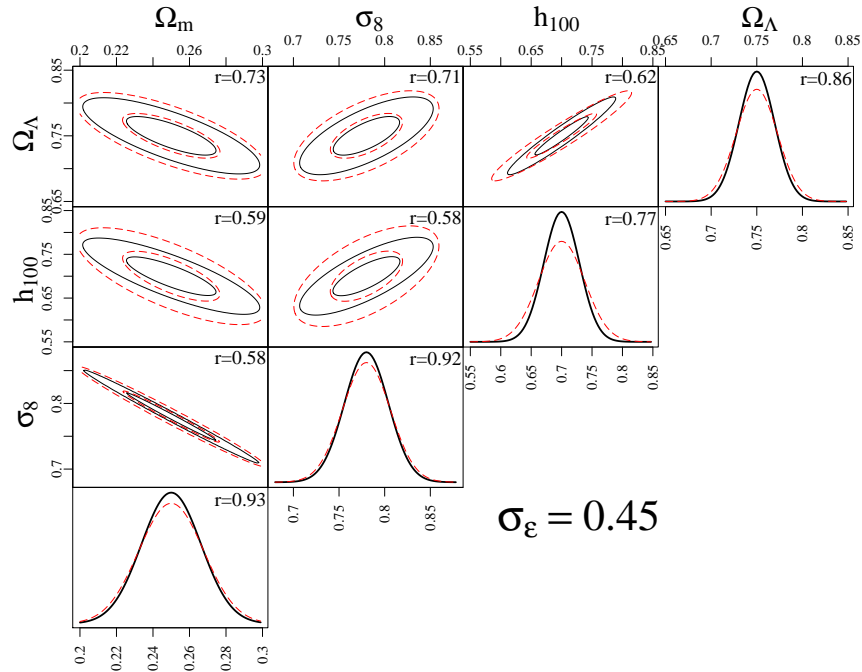
Figure 6.12: *Same as Fig. 6.11, but including shape noise ($\sigma_\epsilon = 0.45$). Note that the scales of the axis are different from those in Fig. 6.11.*

to those of the log-likelihood. We can then compute the Fisher matrix by averaging over all realizations:

$$\hat{\mathsf{F}}_{\alpha\beta} = \frac{1}{N} \sum_{k=1}^{N} \frac{\partial \log p(\boldsymbol{\pi}|\boldsymbol{\xi}^{(k)})}{\partial \pi_\alpha} \frac{\partial \log p(\boldsymbol{\pi}|\boldsymbol{\xi}^{(k)})}{\partial \pi_\beta} \quad , \tag{6.63}$$

where $\boldsymbol{\xi}^{(k)}$ is the correlation function estimated from the k-th realization. In App. D, we show that the expression for the Fisher matrix of the ICA likelihood can be evaluated further to be

$$\mathsf{F}_{\alpha\beta} = \sum_i \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_i}{\partial \pi_\beta} \int \mathrm{d}s_i \, p_{s_i}(s_i) \left(\frac{\partial \log p_{s_i}(s_i)}{\partial s_i} \right)^2 \quad . \tag{6.64}$$

This equation allows a simpler, alternative computation of F from the estimated $p_{s_i}(s_i)$, although it is not free of numerical difficulties, as discussed in App. D.

We use Eqns. (6.62) and (6.63) to compute the Fisher matrices for a 1500-deg^2 survey ($N_\mathrm{f} = 6000$) and find an excellent agreement of the two methods. We fit for four cosmological parameters ($\Omega_\mathrm{m}, \sigma_8, h_{100}, \Omega_\Lambda$), keeping all other parameters fixed to their true values. To visualize the posterior, we compute two-dimensional marginalized posterior distributions for each parameter pair as well as the one-dimensional marginals for each parameter. The results are shown in Fig. 6.11 for a hypothetical survey without shape noise and in Fig. 6.12 including shape noise with $\sigma_\epsilon = 0.45$. A general feature of the ICA likelihood, which has already been apparent in the 2D-analysis (Fig. 6.10), is that the credible intervals are significantly smaller than the ones derived from the Gaussian likelihood. For the two-dimensional marginal distributions, the area of the 68% credible regions derived from the ICA likelihood are smaller by $\approx 30-40\%$. The one-dimensional constrains are tighter by $\approx 10-25\%$. In addition we find that the ICA Fisher ellipses in some cases are slightly tilted with respect to those computed using the Gaussian likelihood. This is particularly apparent for the case without shape noise and for parameter combinations involving the Hubble parameter. Note that the shift of the maximum observed in the two-dimensional case for a single CDFS-like field does is absent here because it was assumed for the Fisher analysis that the posterior is centred on the true parameter values.

We conclude that the non-Gaussianity of the likelihood has a considerable effect on the constraints on cosmological parameters derived from cosmic shear measurements. Although our method is only approximate in that it assumes that a linear transformation suffices to remove statistical dependencies from the data vectors, the statement that the credible regions are smaller than for a Gaussian likelihood seems secure (and is confirmed by our results using the PPDE likelihood). Currently, considerable effort is dedicated to properly account for non-linear contributions to the covariance matrix of cosmic shear (e.g. Takada & Jain 2008; Cooray & Hu 2001b), which increases the size of the error bars on cosmological parameters compared to using a covariance derived for a Gaussian shear field. Our findings suggest that the accurate knowledge of the covariance matrix does not necessarily lead to better error estimates if the non-Gaussian nature of the likelihood is not taken into account.

6.4 How odd is the Chandra Deep Field South?

In their cosmic shear analysis of the combined HST GEMS and GOODS data of the Chandra Deep Field South, Schrabback et al. (2007) (S07 from hereon) have found a very low value of $\sigma_8(\Omega_\mathrm{m} = 0.3) = 0.52^{+0.11}_{-0.15}$. We refer the reader to the original publication for details on
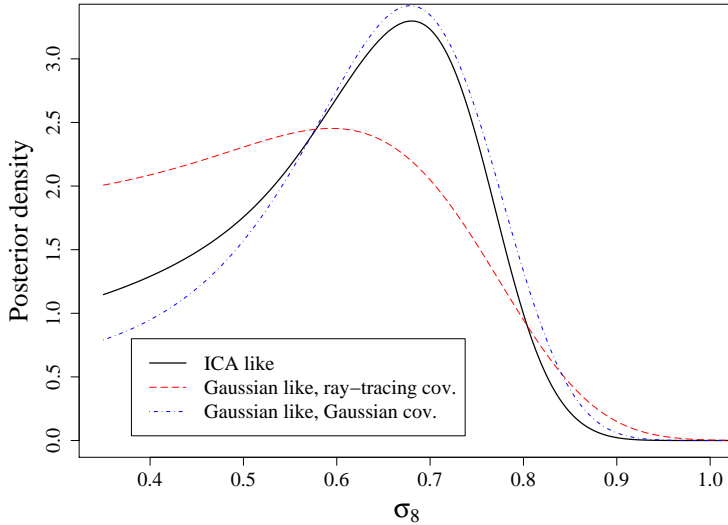
Figure 6.13: *Posterior distributions for σ_8 as computed from the CDFS data. The solid line corresponds to the ICA likelihood, the dashed line is from the Gaussian likelihood whose covariance matrix was estimated from the ray-tracing simulations. The dot-dashed line was computed from the Gaussian likelihood with an analytically computed covariance matrix, assuming that the shear field is Gaussian.*

the data and weak lensing analysis. Here, we present a re-analysis of the cosmic shear data based on a revised source catalogue (work by T. Schrabback), taking into account a previously undetected bias in the shear recovery process. Using our estimate of the non-Gaussian likelihood, we investigate whether cosmic variance alone is responsible for producing the low σ_8-estimate or whether the criteria applied by Giacconi et al. (2001) to select a field suitable for deep X-ray observations have a share in this.

6.4.1 Cosmic Shear analysis of the CDFS

In Fig. 6.13, we show the posterior distribution for σ_8 based on the revised source catalogue. For the fit, all other cosmological parameters were held fixed at the fiducial values chosen for our ray-tracing simulations. We choose a flat prior for σ_8, with a lower boundary of $\sigma_{8,\,\mathrm{min}} = 0.35$ to cut off the tail of the posterior distribution towards small values of the power spectrum normalization, which is caused by the fact that the difference (and therefore the likelihood) between the data and the model vectors changes only very little when σ_8 (and therefore the shear correlation function) is very small. We have performed the fit for the ICA likelihood as well as for the Gaussian approximation to the likelihood. For the latter, the covariance matrix was in one case estimated from the full sample of our ray-tracing simulations, and in the other case computed analytically assuming that the shear field is a Gaussian random field (Joachimi et al. 2008). The striking similarity of the posterior densities derived from the ICA likelihood and using the Gaussian covariance matrix for this particular data vector is merely a coincidence and is in general not seen for our set of simulated correlation functions.

For estimates of σ_8, we use the maximum of the posterior (henceforth we write ICA-MAP for the maximum of the non-Gaussian likelihood, and Gauss-MAP if the Gaussian approximation is used), although we also quote the median (ICA median) for comparison with S07. In the first case, our $\alpha\%$ credible intervals are highest posterior density intervals, whereas for the median we choose to report the interval for which the probability of σ_8 of being below lower interval boundary is as high as being above the upper boundary. The results are summarized in Tab. 6.1.

6.4.2 Cosmic Variance

The original estimates for σ_8 given in S07 and those found in the previous section for the Gaussian likelihood are rather low compared to the value reported by the WMAP5 team (Dunkley et al. 2008). This problem appears less severe when the full non-Gaussian likelihood is used, but the σ_8-estimate is still rather low. It is therefore interesting to know whether this can be fully attributed to cosmic variance or whether the way in which the CDFS was originally selected biases our estimates low.

To begin, we determine the probability of finding a low σ_8 in a CDFS-like field when the pointing is completely random. We estimate the sampling distribution of the σ_8-MAP estimators for Gaussian and ICA likelihoods from the full sample of our ray-tracing simulations. We compute the posterior likelihood for σ_8 using a uniform prior in the range $\sigma_8 \in [0.35; 1.8]$ and determine the MAP estimator $\hat{\sigma}_8$. As in the previous sections, we do this using both the Gaussian and the ICA likelihoods. To separate possible biases of the estimators from biases that might arise because the model prediction based on Smith et al. (2003) does not quite fit our simulations, we correct the simulated correlation functions for this: if $\boldsymbol{\xi}^{(i)}$ is the correlation function measured in the i-th realization, then

$$\boldsymbol{\xi}_{\text{rc}}^{(i)} = \boldsymbol{\xi}^{(i)} - \langle\boldsymbol{\xi}\rangle + \boldsymbol{m}(\boldsymbol{\pi}_0) \,, \tag{6.65}$$

is the "re-centred" shear correlation, where $\langle\boldsymbol{\xi}\rangle$ is the mean of all realizations and $\boldsymbol{m}(\boldsymbol{\pi}_0)$ is our fiducial model.

The resulting sampling distributions of $\hat{\sigma}_8$ are shown in Figs. 6.14 (original ξ) and 6.15 (re-centred ξ). Remarkably, all the distributions are well fit by a Gaussian. With the original correlation functions, we obtain estimates $\hat{\sigma}_8$ which are too high on average. This reflects the fact that the power spectrum fitting formula by Smith et al. (2003) underpredicts the small scale power in the simulations (see Sec. 5.2.1). If we correct for this, we see that the maximum of the ICA likelihood is a nearly unbiased estimator of σ_8 in the one-dimensional case considered here, and in addition has a lower variance than the maximum of the Gaussian likelihood.

We estimate the probability of obtaining a power spectrum normalization as low as the one measured in the CDFS or lower, $\text{Prob}(\hat{\sigma}_8 < \hat{\sigma}_8^{\text{CDFS}})$, by the ratio of the number of realizations which fulfil this condition to the total number of simulations. These estimates agree very well with those computed from the best fitting Gaussian distribution. The results for the MAP and median estimators are summarized in Tab. 6.2. As expected from the above considerations, we find higher probabilities for the re-centred correlation functions. In this case, the ICA-MAP estimator yields 13% for the probability of obtaining an equally low or lower σ_8 than the CDFS. This reduces to $\approx 5\%$ when the uncorrected correlation functions are used, because the misfit of our theoretical correlation functions to the simulations biases the σ_8-estimates high. If we assume that our simulations are a reasonable representation of the real Universe, we can expect the same bias when we perform fits to real data. Therefore, $\text{Prob}(\hat{\sigma}_8 < \hat{\sigma}_8^{\text{CDFS}}) \approx 0.05$ as derived from the uncorrected correlation functions is most likely closest to reality. The probabilities computed from the Gauss-MAP estimates are generally smaller than the ICA-MAP values because of the lower value of $\hat{\sigma}_8^{\text{CDFS}}$ found using these estimators, even though the sampling distributions of the Gauss estimators are broader.
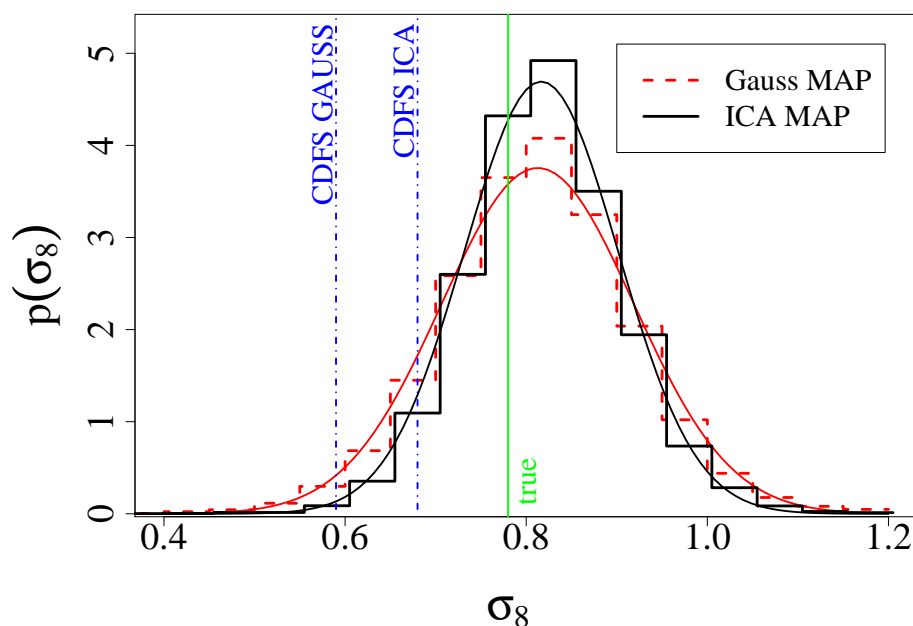
Figure 6.14: *Sampling distributions of the MAP estimators of σ_8, derived from 9600 realizations of the CDFS. All other parameters were held fixed at their fiducial values for the fit. The histogram with red dashed lines has been obtained from the Gaussian likelihood, the one with solid lines from the ICA likelihood. Also shown are the best fitting Gaussian distributions. We indicate the fiducial value of σ_8 and our estimates from the CDFS with vertical lines.*
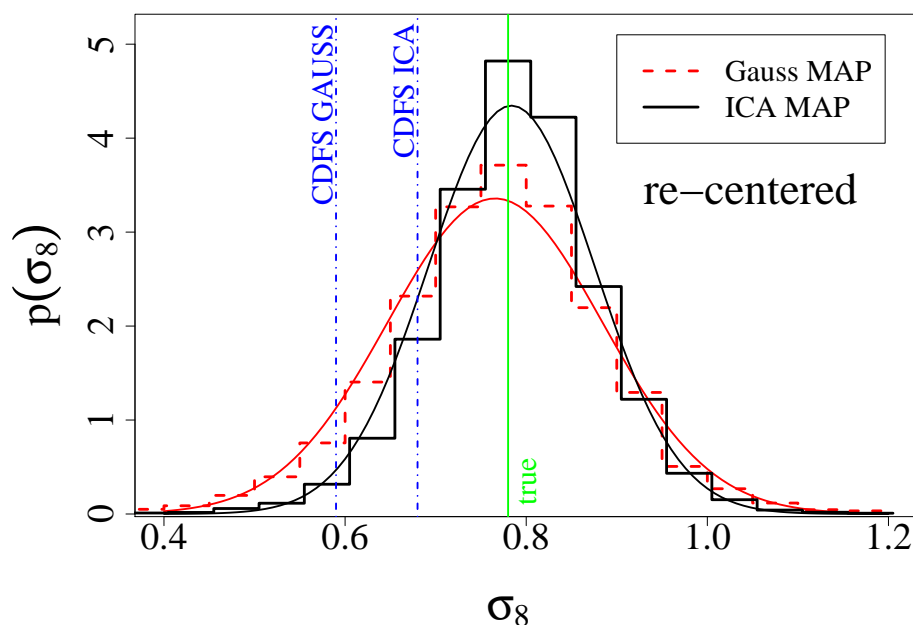

Figure 6.15: *Same as Fig. 6.14, but using re-centred correlation functions*

	ICA likel.	Gaussian likel. (ray-tracing cov.)	Gaussian likel. (Gaussian cov.)
MAP	$0.68^{+0.09}_{-0.16}$	$0.59^{+0.10}_{-0.19}$	$0.68^{+0.10}_{-0.14}$
Median	$0.62^{+0.11}_{-0.11}$	$0.57^{+0.15}_{-0.15}$	$0.64^{+0.10}_{-0.14}$

Table 6.1: *Estimates of σ_8 from the CDFS*

	Gauss (MAP)	Gauss (median)	ICA (MAP)	ICA (median)
re-centred CF	6.8%	8.6%	12.9%	9.0%
original CF	1.8%	3.0%	5.4%	3.4%

Table 6.2: $\mathrm{Prob}(\sigma_8 < \hat{\sigma}_8^{\mathrm{CDFS}})$ *for the CDFS*

6.4.3 Influence of the CDFS selection criteria

We now investigate if and by how much the way in which the CDFS was selected can bias our estimates of the power spectrum normalization low. Several local criteria had to be fulfilled by the future CDFS, such as a low galactic HI density, the absence of bright stars and observability from certain observatory sites. Since these conditions do not reach beyond our galaxy, we do not expect them to affect the lensing signal by the cosmological large-scale structure.

Furthermore, the field was chosen such that no extended X-Ray sources from the ROSAT All-Sky Survey (RASS), in particular galaxy clusters, are in the field of view. This is potentially important, since it is known from halo-model calculations that the cosmic shear power spectrum on intermediate and small scales is dominated by group- and cluster-sized halos. Therefore, the exclusion of X-ray clusters might bias the selection of a suitable line of sight towards underdense fields. On the other hand, the RASS is quite shallow and thus only contains very luminous, nearby clusters, which have a limited impact on the lensing signal due to their low number and low lensing efficiency. We quantify the importance of this criterion using the halo catalogues of our N-body-simulations. To each halo, we assign an X-ray luminosity in the energy range from 0.1 to 2.4 keV using the mass-luminosity relation given in Reiprich & Böhringer (2002) and convert this into X-ray flux using the halo redshift. We then compute the average of the σ_8 estimates from all fields which do not contain a cluster brighter than a certain flux limit. It is difficult to define an exact overall flux limit to describe the CDFS selection, because the RASS is rather heterogeneous. However, it is apparent from Fig. 6.16 that even a very conservative limit of 10^{-13} ergs/sec/cm^2 will change the average σ_8 estimate by at most $3 - 5\%$. This bias is therefore most likely not large enough to explain our CDFS result alone.

Finally, the CDFS candidate should not contain any "relevant NED source". This is very hard to translate into a quantitative criterion, in particular because our simulations contain only dark matter. We model the effect of imposing this requirement by demanding that there be less than n_{g} group- or cluster-sized halos ($M > 10^{13}\,h^{-1}\,M_\odot$) in the redshift range from $z = 0$ to $z = 0.5$ in a CDFS candidate. The impact of this criterion on the estimated value of $\hat{\sigma}_8$ using the ICA- and Gauss-MAP estimators is shown in Fig. 6.17. As expected, the median $\hat{\sigma}_8$ is a monotonically increasing function of n_{g}. For fields with less than ≈ 12 massive halos, the probability of obtaining a power spectrum normalization as low as in the CDFS rises above $\approx 20\%$. Given that the average number of massive halos in the specified redshift range is 18.5, it does not seem to be too unreasonable that fields with less than ≈ 12 such halos could be
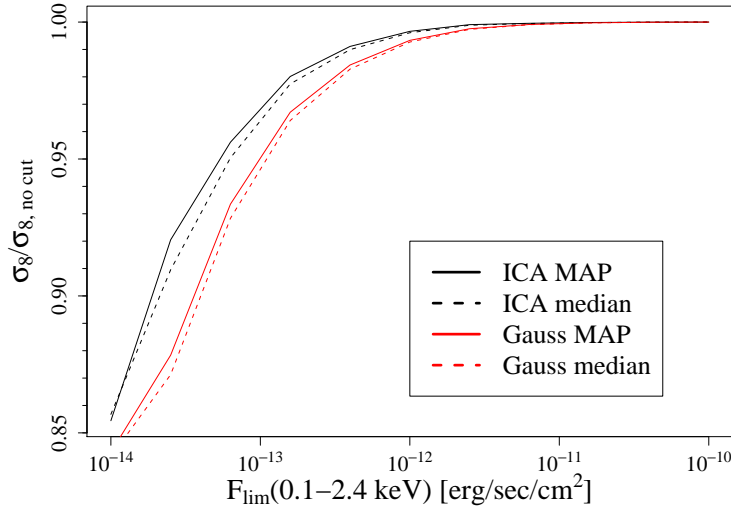
Figure 6.16: *The average values of the ICA-MAP (solid black line) and Gauss-MAP (solid red line) estimators computed from CDFS realizations that do not contain clusters with an X-ray flux larger than F_{lim}. For comparison, we also plot the averages of the corresponding median estimators (dashed lines).*

obtained by selecting "empty" regions in the NED. From Fig. 6.17, we therefore estimate that this selection criterion can cause a bias of σ_8 of another 5-10%.

Note that the two criteria discussed in this section are not strictly independent. However, it is highly improbable that a single field will contain more than one massive halo that would exceed the X-ray flux limit. Therefore, selecting fields without an X-ray-bright cluster prior to performing the steps that lead to Fig. 6.17 would change the halo numbers that go into the analysis by at most one and would not significantly influence our estimate of the bias.

We include these biases as a systematic error of 15% in the error budget of our estimate of the power spectrum normalization from the CDFS, which then reads $\sigma_8 = 0.68^{+0.09(\text{stat.}) \, +0.1(\text{sys.})}_{-0.16(\text{stat.})}$.
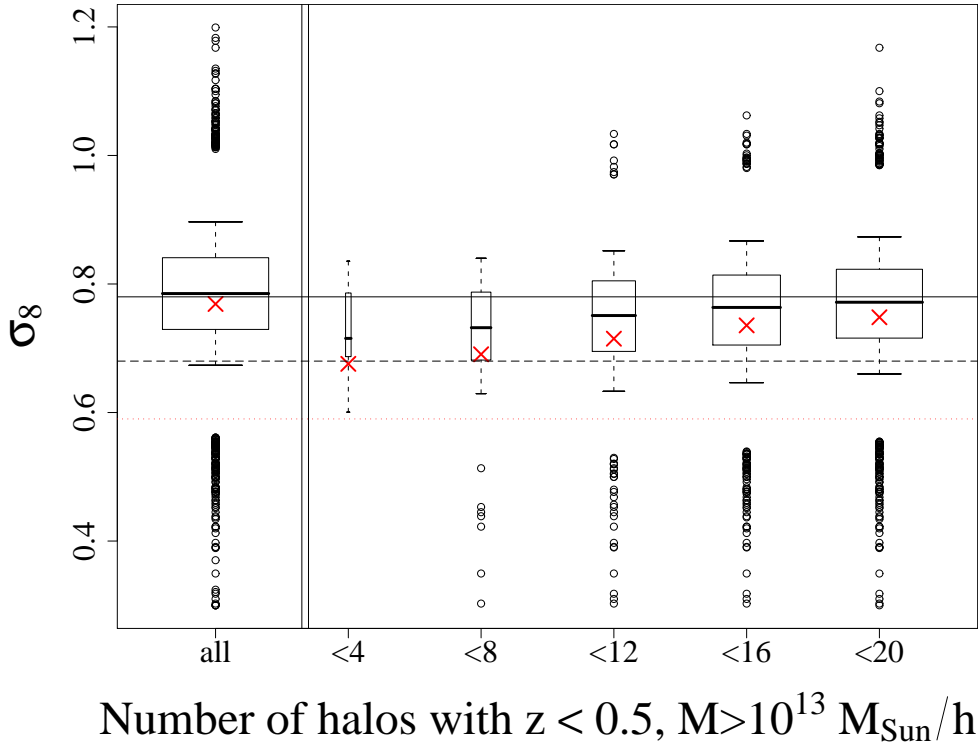
Figure 6.17: *Dependence of the ICA-MAP-estimator for σ_8 on the number of group- and cluster-sized halos n between $z = 0$ and $z = 0.5$. For each n-bin, we summaries the distribution of the corresponding subsample of simulated CDFS-fields by giving a box plot: The thick horizontal line in each box denotes the median, the upper and lower box boundaries give the upper and lower quartiles of the distribution of the sample values. The error bars ("whiskers") extend to the 10% and 90% quantiles, respectively. The values outside of the 5% and 95% quantiles are given as points. The width of each box is proportional to the square root of the sample size. For comparison, we also show for each subsample the median of the Gauss MAP estimators as red crosses. The solid black horizontal line indicates the true value of σ_8, the black dashed line the ICA-MAP estimate for the CDFS and the red dotted line the Gauss-MAP estimate. The average number of halos with $M > 10^{13}\,h^{-1}\,M_\odot$ and $z \leq 0.5$ in a CDFS-like field is $\bar{n} = 18.5$.*

7 Galaxy-Galaxy-Lensing with the Millennium Simulation

Much work is currently dedicated to improving the understanding of the formation and evolution of galaxies within the cosmological standard model. The general picture is widely agreed on: at early times, baryonic matter essentially follows the dark matter distribution until the gas density is high enough for the gas to cool efficiently. The baryons then condense at the centers of the dark matter halos and there form stars and galaxies. The details of this picture, such as the conditions under which star formation occurs and the role of feedback from quasar activity and supernovae, however, are still little understood. The importance of these effects on the evolution of a galaxy strongly depends on the properties of the dark matter halo hosting the galaxy. Galaxy-galaxy lensing offers the opportunity to link the observed properties of the lens galaxies to their dark matter environment and therefore can be used to put constraints on galaxy evolution models such as the semi-analytic models discussed in Sec. 4.4.

In this Chapter, we apply the ray-tracing code that we have developed for the Millennium Simulation to gain a basic understanding of what information is encoded in the GGL signal. We make predictions for realistic lens samples and compare our results to the measurements of Sheldon et al. (2004) in the SDSS. Previous studies of this kind have been undertaken by Guzik & Seljak (2001) and Yang et al. (2003), who used the semi-analytic model by Kauffmann et al. (1999), or Tasitsiomi et al. (2004), who do not use a full galaxy evolution model but assign a luminosity to halo central particles based on the maximum circular velocity of a halo. Our work constitutes a significant improvement over these previous efforts, not only because we are using the Millennium Simulation with its unprecedented dynamic range, but also because for the first time we combine a complete ray-tracing with a semi-analytic model of galaxy formation.

The results presented in this Chapter are derived from the Millennium Simulation and the semi-analytic model of galaxy formation by De Lucia & Blaizot (2007) throughout. Unless stated otherwise, we use 128 realizations of a $2° \times 2°$ field, for which we have performed ray-tracing up to $z = 1$.

7.1 Galaxy-galaxy lensing as function of the lens sample

7.1.1 Contributions to the galaxy-galaxy lensing signal

The GGL signal can be used to infer the average properties of the dark matter halos hosting the lens galaxies (e.g. Mandelbaum et al. 2008; Johnston et al. 2007; Limousin et al. 2005; Wilson et al. 2001; Schneider & Rix 1997). However, most of these halos are not isolated; they are likely to be embedded into galaxy groups and clusters, which will likewise contribute to the tangential shear profile with respect to the lenses. To understand the effects of the environment of the lens halo, we first study the GGL signal for lens samples selected according to mass and environment before turning to more realistic situations.

In Fig. 7.1, we plot the excess surface mass density $\Delta\Sigma$ (see Eq. 3.111) around a lens as a function of the projected proper separation R from the lens, varying the mass $M_{\rm g}$ of the galaxy
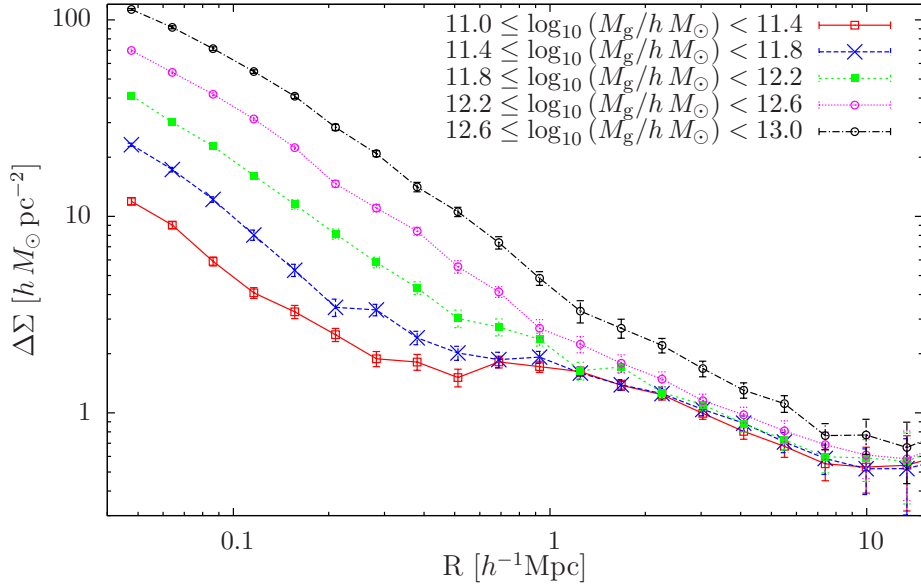
Figure 7.1: *Galaxy-galaxy lensing signal for lens samples selected according to the mass of the subhalo containing the lens galaxy. Only galaxies with $z < 0.4$ were included. The background galaxies are located at $z = 0.9$. For better visibility, the data points have been connected with lines. The errors are estimated from the field-to-field variance, no shape noise was included.*
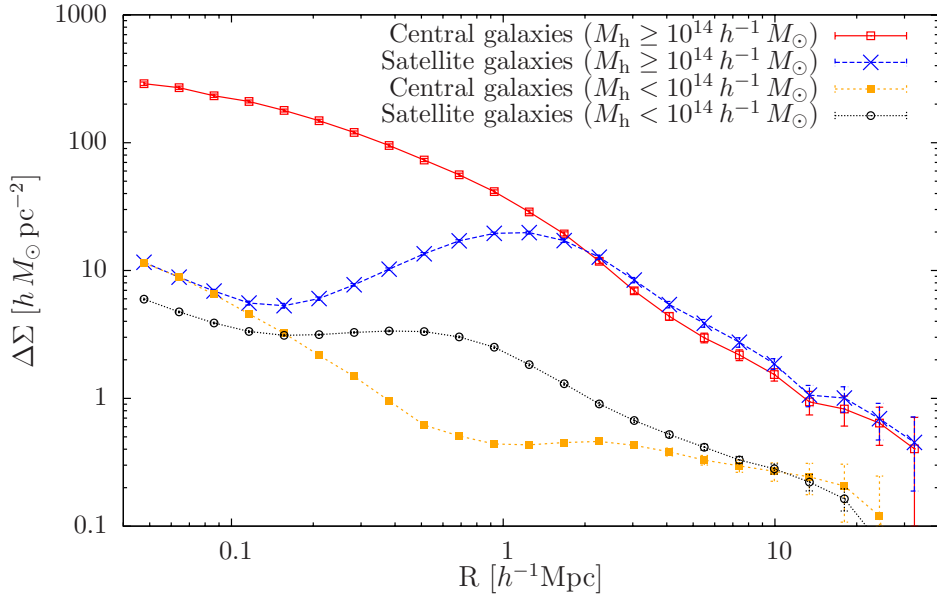

Figure 7.2: *Galaxy-galaxy lensing signal for central galaxies of massive halos with $M_h > 10^{14}\ M_\odot$, galaxies in halos with $M > 10^{14}\ M_\odot$ that are not central galaxies, central galaxies of halos with $M_h \leq 10^{14}\ M_\odot$ and galaxies in halos with $M_h \leq 10^{14}\ M_\odot$ which are not central galaxies. Galaxies with $z < 0.9$ were included in the samples. The background galaxies are all located at $z = 0.9$. For better visibility, the data points have been connected with lines. The errors are estimated from the field-to-field variance, no shape noise was included.*

halos, i.e. of the (sub-)halos the lens galaxies reside in. We use five mass bins in the range $10^{11}\,M_\odot/h \leq M_\mathrm{g} < 10^{13}\,M_\odot/h$, where for each bin $\Delta M/M = 0.4$. On small scales, the GGL signal increases roughly proportional to M_g. However, $\Delta\Sigma$ displays a clear break on larger scales, beyond which the signal is significantly flatter and depends only weakly on the mass of the lens halo. The location of this break is seen at smaller separations for lens halos with lower mass. For the lowest-mass bins, even a secondary maximum becomes visible. Clearly, at these scales the GGL signal is no longer dominated by the dark matter halo of the lens galaxy but by its large-scale environment.

This becomes particularly apparent if we select lens galaxies that are satellite galaxies of massive host halos ($M_\mathrm{h} \geq 10^{14}\,M_\odot/h$), as is shown in Fig. 7.2 with the blue dashed curve. In this case, the secondary maximum becomes the dominant feature of the GGL signal, only on very small scales (below $\approx 100\,h^{-1}\,\mathrm{kpc}$) the actual lens halo is important. On large scales, on the other hand, $\Delta\Sigma$ becomes very similar to the signal obtained from the central galaxies of the massive host halos (solid red curve): thus, if R is much larger than the separation between lens and host halo center, $\Delta\Sigma$ essentially probes the shear profile of the host halo. The same effect is seen – with lower amplitude – for the satellite galaxies of less massive host halos ($M_\mathrm{h} < 10^{14}\,M_\odot/h$, dotted black curve). The signal from the central galaxies of these halos (short dashed orange curve), however, now also displays a pronounced break at $\approx 1\,h^{-1}\,\mathrm{Mpc}$. Since we are already looking at central galaxies, this can not be interpreted as the contribution of a possible host halo. Rather, this has to be caused by other correlated structures in the vicinity of the lens galaxy halo which are not well described with the simple two-halo description employed above for the more massive halos.

7.1.2 Modelling the galaxy-galaxy lensing signal

The observations made in the previous section motivate a simple model, in which we interpret the galaxy-galaxy lensing signal as being composed of contributions by the lens galaxy halo, a possible group or cluster environment (the host halo) and projected large-scale structure that is correlated with the lenses. Similar or related models have been proposed in Johnston et al. (2007) and Yang et al. (2006). Both works mainly aim to model the shear signal of galaxy groups and clusters. Our model differs from these works in that we compute the host halo contribution as an average over all possible host halo positions, whereas in Johnston et al. (2007) only central galaxies are considered and Yang et al. (2006) assume fixed separations between subhalos and host halos.

The model

To model the mass distribution associated with the lens galaxies, we assume that each galaxy is located at the center of a dark matter halo. This halo can be an isolated halo, or a subhalo of a more massive host halo. The average surface mass density associated with a lens as function of the projected proper separation R is then composed of three different contributions:

$$\Sigma(R; z) = \Sigma_\mathrm{g}(R; z) + (1 - p_\mathrm{c})\Sigma_\mathrm{h}(R; z) + \Sigma_\mathrm{LSS}(R; z) \,. \tag{7.1}$$

Here, p_c is the probability that the lens galaxy is the central galaxy of an isolated halo (as opposed to being a galaxy at the center of a subhalo). The surface mass density profile of an isolated halo or subhalo with the galaxy at its center is given by Σ_g. In the case that the galaxy halo is embedded in a larger host halo (which occurs for a fraction of $1 - p_\mathrm{c}$ of all lens galaxies),

the host halo contribution is denoted by $\Sigma_{\mathrm{h}}(R)$. Finally, Σ_{LSS} accounts for the projection of additional large-scale structure that is correlated with the lens. For now, we assume that all lens galaxies are located at the same redshift z. The GGL signal for a lens population that extends over a certain redshift range can then be obtained by averaging $\Delta\Sigma(R; z)$ over the redshift distribution of the lens galaxies.

In order to compute Σ_{g} and Σ_{h}, we have to specify the density profiles of the dark matter halos involved. We will employ the *NFW profile*, which has been found (Navarro et al. 1997) to provide a good match to the mean density profile of dark matter halos for a wide range of cosmologies.

The NFW profile

The three-dimensional NFW density profile $\rho(r)$ is given by

$$\rho_{\mathrm{NFW}}(r; z) = \frac{\delta_{\mathrm{c}}\, \rho_{\mathrm{crit}}(z)}{r/r_{\mathrm{s}}(1 + r/r_{\mathrm{s}})^2} \; . \tag{7.2}$$

Here, δ_{c} is a characteristic density contrast and $\rho_{\mathrm{crit}}(z)$ is the critical density of the Universe at redshift z (Eq. 2.10). For small radii r, the density decreases $\propto r^{-1}$, whereas beyond the *scale radius* r_{s} the profile steepens, so that $\rho \propto r^{-3}$. Instead of parametrizing the density profile with the two free parameters r_{s} and δ_{c}, it is more convenient to use the *virial mass* M_{200} within the *virial radius* r_{200}, and the *concentration parameter* c, which is defined by $c = r_{200}/r_{\mathrm{s}}$. Note that several similar parametrizations of the NFW profile with different definitions of the virial radius exist in the literature. In choosing r_{200} to be the virial radius, we adopt the conventions used in the Millennium Simulation database. The virial radius r_{200} is defined as the radius inside which the mean mass density of the halo is $200\rho_{\mathrm{crit}}$, so that the virial mass can be computed using

$$M_{200} = \frac{800\pi}{3}\rho_{\mathrm{crit}}(z)\, r_{200}^3 \; . \tag{7.3}$$

The mass within radius r is given by

$$M(< r) = \frac{4\pi}{c^3}\, \delta_{\mathrm{c}}\, \rho_{\mathrm{crit}}(z)\, \left[\frac{1}{1 + cx} + \log\left(1 + cx\right) - 1\right], \qquad \text{where } x = \frac{r}{r_{200}} \; . \tag{7.4}$$

From this, one can derive a relation between the characteristic density contrast in Eq. (7.2) and the concentration parameter:

$$\delta_{\mathrm{c}} = \frac{200}{3}\frac{c^3}{\ln(1 + c) - c/(1 + c)} \tag{7.5}$$

by demanding that $3M(< r_{200})/(4\pi r_{200}^3) = 200\rho_{\mathrm{crit}}$. An undesirable consequence of Eq. (7.4) is that the mass of a NFW halo for $r \to \infty$ is infinite. For this reason, we use a *truncated NFW profile* of the form (e.g. Baltz et al. 2007)

$$\rho_{\mathrm{NFW,t}}(r; z) = \frac{\delta_{\mathrm{c}}\, \rho_{\mathrm{crit}}(z)}{r/r_{\mathrm{s}}(1 + r/r_{\mathrm{s}})^2}\frac{1}{1 + (r/r_{\mathrm{t}})^2} \; , \tag{7.6}$$

where r_{t} is the *truncation radius*. We choose $r_{\mathrm{t}} = 3r_{200}$, which roughly corresponds to the radius at which the halo density is comparable to the mean density.

Contributions to the surface mass density

We now discuss the individual contributions to our model for $\Delta\Sigma(R)$ in detail. We proceed in two steps: we first compute the contributions to the projected surface mass density $\Sigma(R)$, from which then $\Delta\Sigma(R)$ can be computed using Eq. (3.111).

We obtain the surface mass density due to a dark matter halo by projecting the three-dimensional density profile according to

$$\Sigma(R) = 2 \int_0^\infty \mathrm{d}z \, \rho\left(\sqrt{R^2 + z^2}\right) \, , \qquad (7.7)$$

where z is the coordinate in the direction of the line of sight. This equation can be immediately applied to obtain the projected density Σ_g due to the galaxy halo.

For a possible host halo, the situation is more complicated: since the position of the host halo with respect to the galaxy halo is different for each lens galaxy, we have to average over all possible host halo locations. The geometry of the configuration of host and subhalo is illustrated in Fig. 7.3. Writing the two-dimensional probability distribution of finding a host halo at proper

Figure 7.3: *Configuration of galaxy and host halos in the plane of the sky.*

separation $\boldsymbol{x}$ from the galaxy halo in the plane of the sky in polar coordinates,

$$p(\boldsymbol{x}) \, \mathrm{d}^2 x = p_R(R_\mathrm{h}) \, R_\mathrm{h} \, \mathrm{d}R_\mathrm{h} \, \frac{\mathrm{d}\varphi}{2\pi} \, , \qquad (7.8)$$

the host halo contribution to Σ is given by (see Fig. 7.3)

$$\Sigma_\mathrm{h}(R) = \int_0^\infty \mathrm{d}R_\mathrm{h} \, R_\mathrm{h} \, p_R(R_\mathrm{h}) \int_0^{2\pi} \frac{\mathrm{d}\varphi}{2\pi} \, \Sigma_\mathrm{host}\left(\sqrt{R^2 + R_\mathrm{h}^2 - 2RR_\mathrm{h}\cos\varphi}\right) \, , \qquad (7.9)$$

where Σ_host is the projected density profile of the host halo according to Eq. (7.7). To obtain p_R, we have measured the two-dimensional, spherically averaged distribution of subhalos with galaxies in the Millennium Simulation at $z = 0$. For that, we selected all halos with $M > 10^{13} \, h^{-1} \, M_\odot$, projected the galaxy distribution onto an arbitrary face of the simulation box and counted the number of satellite galaxies within annuli centered on the host halo. As can be seen from Fig. 7.4, the distribution is almost flat near the center of the host halo and declines quickly beyond the virial radius. Since the distribution is very smooth on the relevant length scales, we directly use it for our model by interpolating between the tabulated measured values.

The final ingredient for our model is the contribution by projected large-scale structure that is not accounted for by the galaxy and host halos. In three dimensions, the average density distribution near a lens galaxy is given by the cross-correlation function of galaxies and dark matter, $\xi_\mathrm{g,DM}(r)$:

$$\rho(r; z) = \bar{\rho}_\mathrm{m}(z) \left[1 + \xi_\mathrm{g,DM}(r)\right] \, , \qquad (7.10)$$

where $\bar{\rho}_\mathrm{m}(z)$ is the mean matter density at redshift z. We relate $\xi_\mathrm{g,DM}$ to the correlation function of the dark matter distribution by $\xi_\mathrm{g,DM} = b \, r \, \xi_\mathrm{DM}$, where b and r are the linear stochastic bias
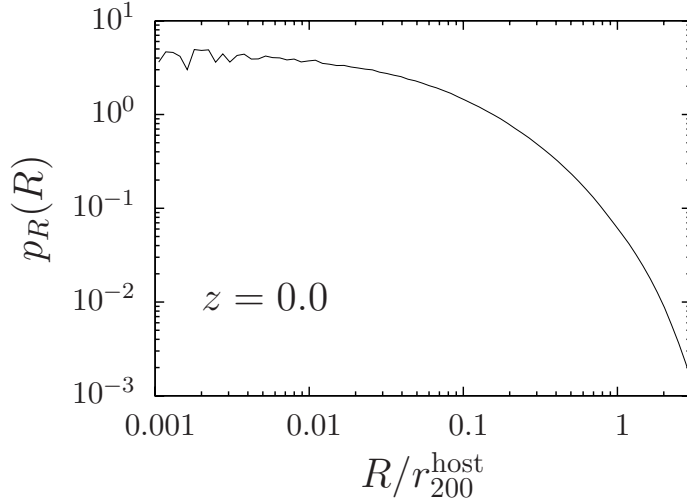
Figure 7.4: *Two-dimensional radial distribution p_R of the galaxies from the model of De Lucia & Blaizot (2007) at $z = 0$ within halos with mass $M > 10^{13} \, h^{-1} \, M_\odot$.*

and correlation factors, respectively (see Chapter 8 for details). We assume that b and r can be approximated to be constant, since the projection of large-scale structure becomes important for our model only on large scales. Furthermore, we approximate ξ_{DM} with the prediction from linear perturbation theory. This is justified because this contribution becomes important only on scales far beyond $1 \, h^{-1} \, \mathrm{Mpc}$, where effects of non-linear structure formation can be neglected. With these approximations, the resulting surface mass density can be found by a projection similar to Eq. (7.7):

$$\Sigma_{\mathrm{LSS}}(R) = b \, r \, \rho_{\mathrm{crit},0} \, \Omega_{\mathrm{m}} \, (1 + z)^3 \times 2 \int_0^\infty \mathrm{d}z \, \left[1 + \xi \left(\sqrt{R^2 + z^2} \right) \right] . \tag{7.11}$$

With this, the surface mass density associated with lens galaxies is fully specified. The next step is to compute the mean mass within a radius R using

$$\bar{\Sigma}(< R) = \frac{2}{R^2} \int_0^R \mathrm{d}R' \, R' \, \Sigma(R') . \tag{7.12}$$

The excess surface mass density is then given by Eq. (3.111):

$$\Delta\Sigma(R) = \bar{\Sigma}(< R) - \Sigma(R) . \tag{7.13}$$

In total, our model has six free parameters: the virial masses M_{g} and M_{h} and concentration parameters c_{g} and c_{h} of the galaxy and the host halo, respectively, the probability p_{c} that a lens galaxy is the central galaxy of an isolated halo, and the product br of the bias and correlation parameters.

The model is approximate in several respects: we do not take into account the scatter in halo masses and concentration parameters, which in realistic lens samples can be quite substantial (e.g. Neto et al. 2007). Since the model depends non-linearly on these parameters, it can not be expected to yield particularly good estimates of the "mean" masses and concentrations if the lens sample is very heterogeneous. Furthermore, we assume that the halo profiles of subhalos are well described by the same NFW profile that is used for the isolated lens halos. However, the subhalo profiles are likely to be heavily truncated by tidal stripping within their host. We therefore can only obtain an average lens halo profile to which these effects are folded in. Finally,

in practice we compute the model prediction for $\Delta\Sigma$ for the mean redshift of the lens galaxies, which we find be accurate to within a few percent for the lens samples considered here. If the lens galaxies extend over a larger redshift range, the model would have to be properly averaged over the lens redshift distribution.

Comparison to the simulation

We have tested the model described in the previous section on various lens galaxy samples, the properties of which are summarized in Tab. 7.1. To minimize the effect of the scatter of the halo mass and the concentration parameters, we first apply the model to the five mass-selected samples M1-M5 (see also Fig. 7.1). We then proceed to fit the model to two more realistic lens samples (called blue and red sample henceforth), which we have selected according to their observer-frame $u - r$ color. Since the galaxies in the semi-analytic model show a clearly bimodal color distribution with a minimum at $u - r = 2.2$, we choose this number to separate red from blue galaxies.

We first obtain a best-fit-solution by minimizing the following χ^2-function with respect to the parameter vector $\boldsymbol{\pi} = (\log M_\mathrm{g}/h\, M_\odot,\, M_\mathrm{h}/h\, M_\odot, c_\mathrm{g}, c_\mathrm{h}, p_\mathrm{c}, br)^\mathrm{t}$:

$$\chi^2(\boldsymbol{\pi}) = \sum_{i,j=1}^{p} \left(\Delta\boldsymbol{\Sigma}(\boldsymbol{\pi}) - \Delta\hat{\boldsymbol{\Sigma}}\right)^\mathrm{t} \mathsf{C}_{ij}^{-1} \left(\Delta\boldsymbol{\Sigma}(\boldsymbol{\pi}) - \Delta\hat{\boldsymbol{\Sigma}}\right) , \tag{7.14}$$

where $\Delta\hat{\boldsymbol{\Sigma}} = (\Delta\hat{\Sigma}(R_1), \ldots, \Delta\hat{\Sigma}(R_p))^\mathrm{t}$ is the estimate of $\Delta\Sigma$ from the simulation in the radial bin centered on R_i, $\Delta\boldsymbol{\Sigma}(\boldsymbol{\pi})$ is the corresponding model prediction and p is the number of radial bins. In writing Eq. (7.14), we have implicitly assumed a Gaussian likelihood function for the GGL signal.

We have estimated the covariance matrix C from the field-to-field variance of our 128 ray-tracing realizations. Since we are using $p = 20$ radial bins, it is important to de-bias the inverse of this matrix using Eq. (6.22). In Figs. 7.5 and 7.6, we show the correlation matrices for the blue and red lens galaxy samples, respectively. The matrices display a block-diagonal structure, each block showing a different correlation structure. The different sub-blocks can be associated with the three contributions to our model: on scales where the lens galaxy halo dominates the GGL signal, the correlation matrices show a narrow, band-diagonal structure. In the red sample, two further sub-blocks can be identified, each corresponding to the ranges over which the host contribution and the large-scale structure term are important. In the blue sample, the host halo term is almost always neglegible, and consequentially the corresponding block is absent in the correlation matrix. Adding shape noise due to the ellipticities of the background galaxies ($\sigma_\epsilon = 0.3$, where σ_ϵ is the dispersion of each ellipticity component) makes the correlation matrix significantly more diagonally dominated.

To obtain the full posterior distribution of our model parameters, we use the parameter estimates from these fits as starting points for Monte-Carlo-Markov-Chains (e.g. Robert & Casella 2005). We derive error estimates for the individual parameter estimates by marginalizing over all other parameters and computing the highest-posterior-density credible regions.

The best-fit parameters can then be compared to the halo properties directly inferred from the halo trees in the Millennium Simulation database. While this is straightforward for the halo masses, for which we compute the mean mass of the halos in the lens galaxy sample, the concentration parameters are not directly provided by the database. One possibility to obtain them is from the ratio of the maximum rotation velocity v_max of a halo to the rotation velocity

at the virial radius, v_{200}. The rotation curve for the NFW profile is given by (Navarro et al. 1997)

$$\frac{v(r)}{v_{200}} = \left(\frac{1}{x} \frac{\log(1+cx) - cx/(1+cx)}{\log(1+c) - c/(1+c)} \right)^{1/2} \ . \tag{7.15}$$

The value of x that maximizes Eq. (7.15) can be determined numerically for a range of values of c, which then yields $v_{\max}/v_{200}$ as a function of the concentration parameter. Unfortunately, it turns out that this function is not invertible even for reasonable values of c. However, we do not have to compute the concentration parameter for every halo individually, since our model only yields halo properties that are averages over the whole halo population in the lens sample. We therefore use the fitting formula for the concentration-mass relation provided by Neto et al. (2007),

$$c(M_{200}, z) = \frac{4.67}{1+z} \left(\frac{M_{200}}{10^{14} \, h^{-1} \, M_\odot} \right)^{-0.11} \tag{7.16}$$

to compute c from the mean halo mass.

The best fitting models for the mass-selected lens samples are shown in Fig. 7.7 and for the color-selected samples in Fig. 7.9. The best-fit parameters are summarized and compared to the "true" average values in the Millennium simulation in Tab. 7.3. In Fig. 7.8, we show as representative examples the two- and one-dimensional marginalized posterior distributions of the model parameters for M1 and the blue sample. We find that our model fits the GGL signal of the samples M1-M5 very well. As expected, the galaxy halo contribution dominates $\Delta\Sigma$ out to larger distances from the lens the higher M_{g}. The mass of the galaxy halo is very accurately predicted by the model, generally with an accuracy of 6% or better. The estimated concentration parameters for the lens galaxy halos are compatible with the predictions from Eq. (7.16) for the samples, albeit with a larger uncertainty of $\approx 15\%$. The properties of the host halos are less well constrained. This is in parts because in most samples $\Delta\Sigma_{\mathrm{h}}$ never comes to fully dominate the GGL signal, in contrast to the lens halo contribution. Furthermore, both host halo mass and concentration parameter are strongly degenerate with the fraction p_{c} of lens galaxies in isolated halos, because an increase the host halo mass can be compensated by reducing the fraction of lens galaxies that are in subhalos. The estimates of br are essentially independent of the other model parameters. Since the large-scale structure contribution becomes important only for large R, where the measurement errors are large due to cosmic variance, this parameter cannot be constrained very accurately.

As expected, the performance of the model is less good for the more heterogeneous blue and red samples (Fig. 7.9 and Tab. 7.3). The model completely fails to reproduce the break in $\Delta\Sigma$ of the blue sample at $\approx 1 \, h^{-1} \, \mathrm{Mpc}$, which the model attributes entirely to large-scale structure. This explains the very weak constraints on the host halo properties seen in Fig. 7.8. For both samples, in particular the estimates of the galaxy halo masses and concentrations are incompatible with the values obtained from the halo catalogs of the Millennium Simulation. Clearly, a more accurate description taking into account the scatter of the parameters characterizing the dark matter halos involved is needed in order to make unbiased statements about the properties of the matter associated with galaxies. A natural candidate to achieve this is the halo occupation distribution (HOD) formulation (e.g. Seljak 2000), which we will test against our simulations in future work.
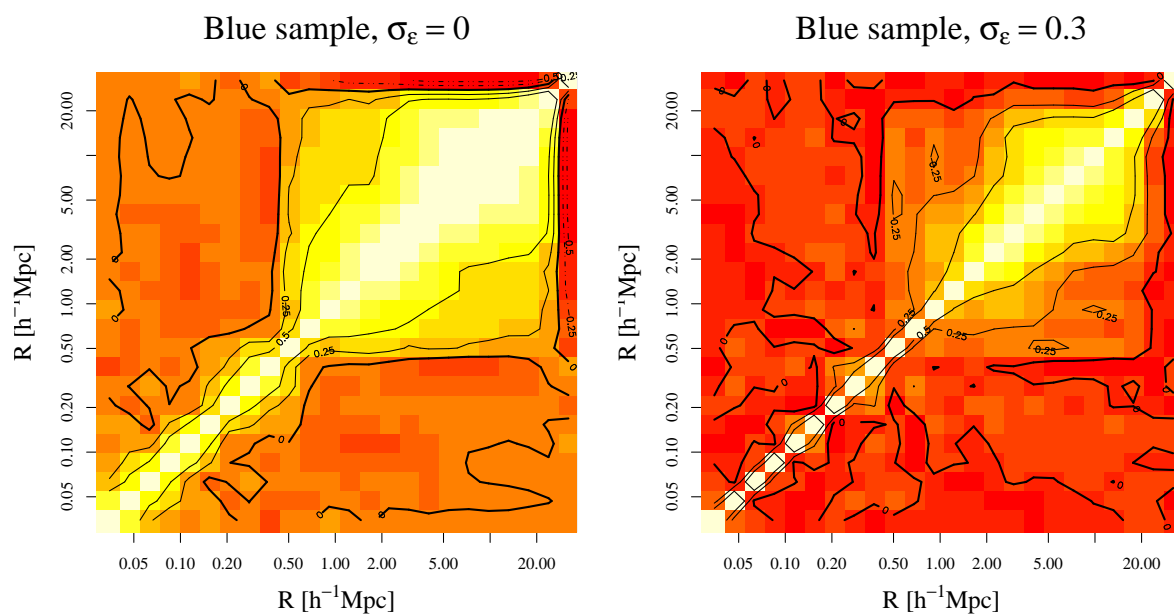
Figure 7.5: *Correlation matrix for the GGL signal of the blue sample without (left panel) and with shape noise (right panel). Light colors indicate high correlation.*
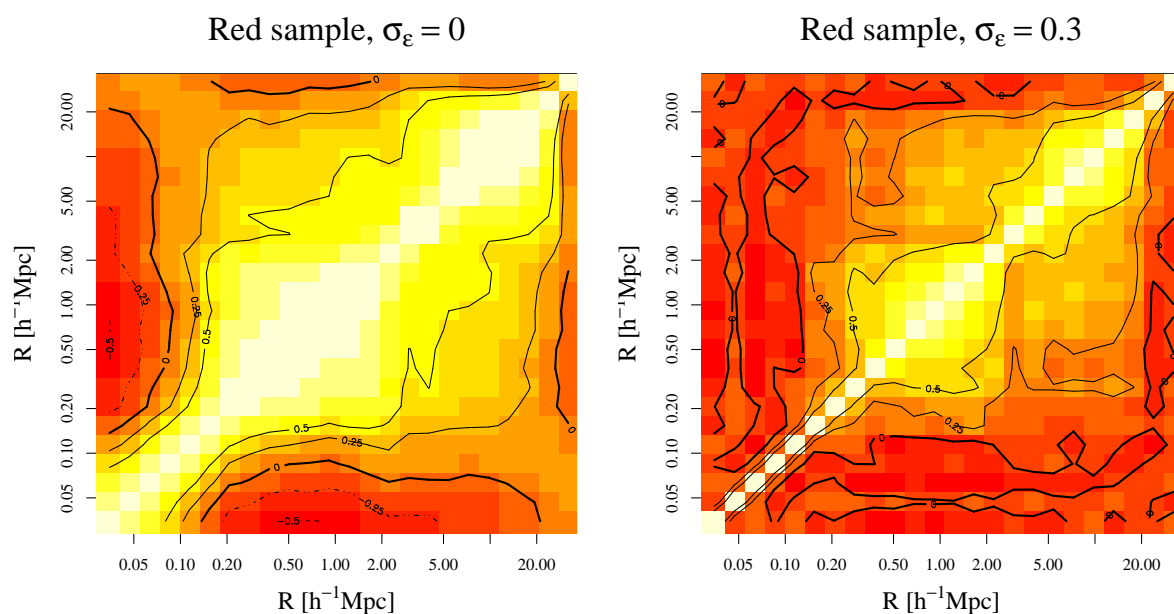

Figure 7.6: *Correlation matrix for the GGL signal of the red sample without (left panel) and with shape noise (right panel). Light colors indicate high correlation. Anti-correlation is signified by the dashed contours.*
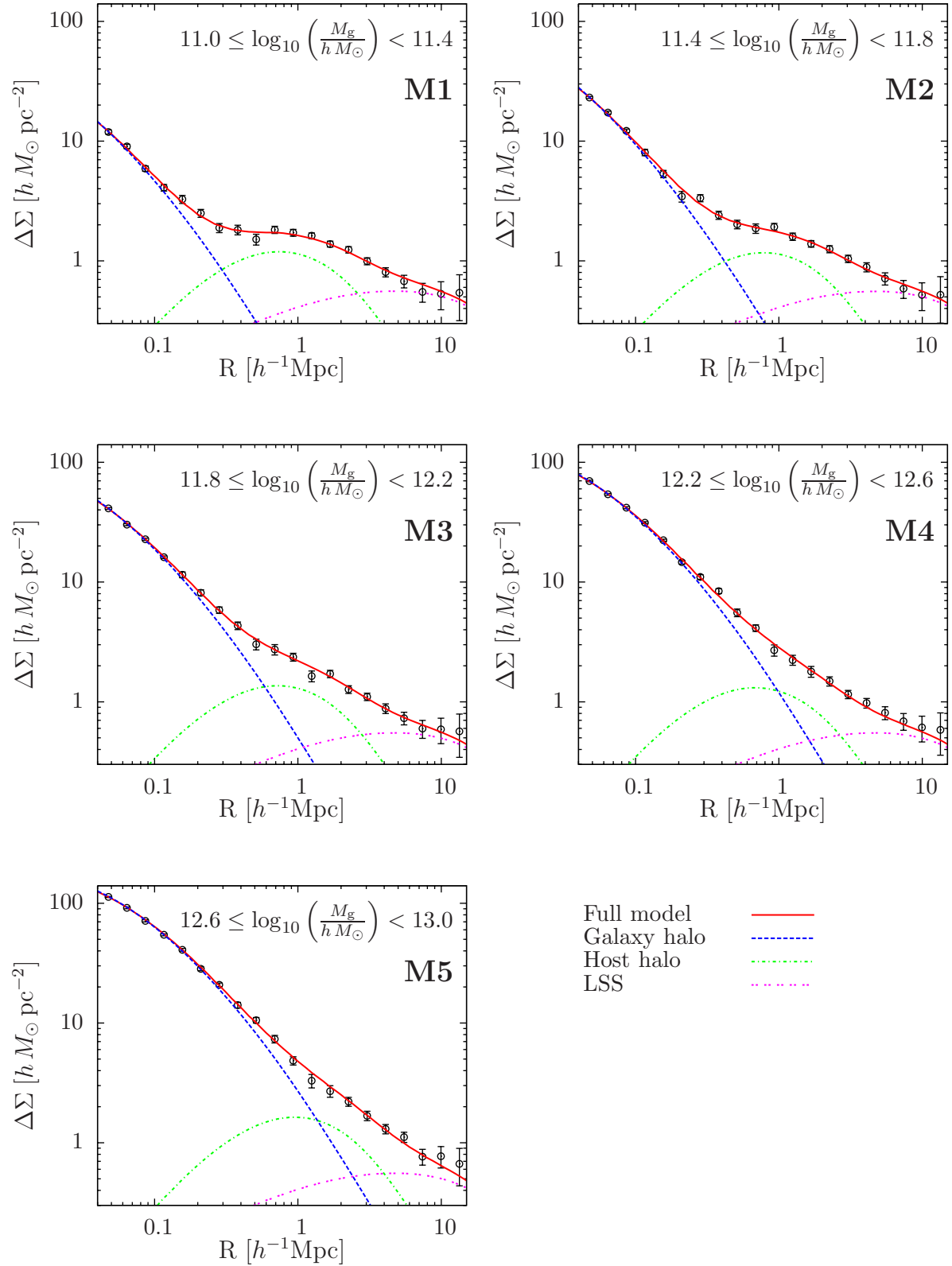
Figure 7.7: *Model fits to the GGL signals of the five mass-selected samples M1-M5 (see Tab. 7.1). The best-fitting model is shown as solid red lines. The individual contributions to the modelled signal are given by dashed blue lines (galaxy halo), dot-dashed green lines (host halo) and wide-dashed magenta lines (large-scale structure).*
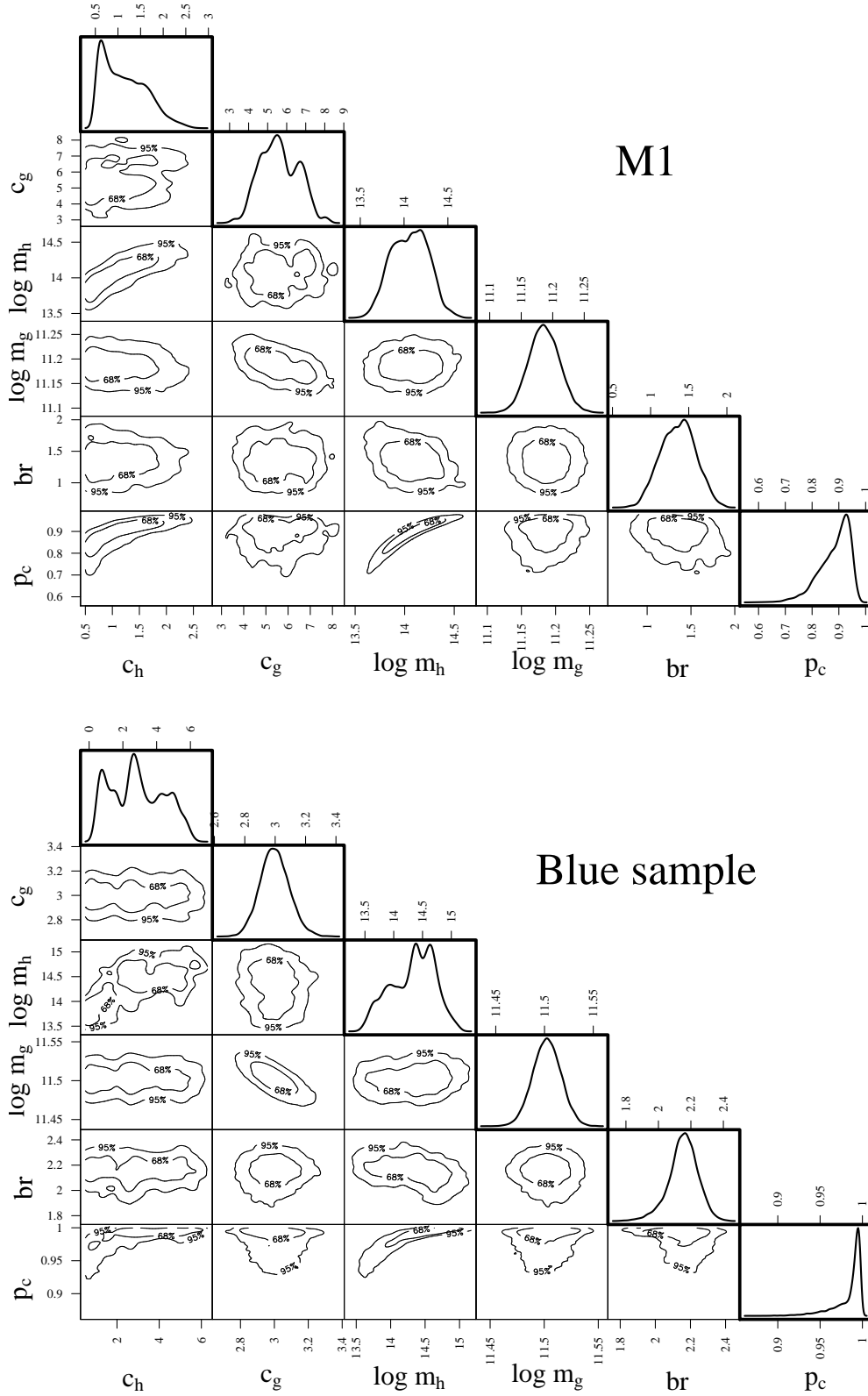
Figure 7.8: *Two- and one-dimensional marginalized posterior distributions (off-diagonal and diagonal panels, respectively) for the M1 and blue lens galaxy samples. In the axis labels, we use the abbreviations* $m_{g,h} = M_{g,h}/(h\,M_\odot)$.

Name	Selection criteria
M1	$11.0 \leq \log_{10} (M_{\mathrm{g}}/hM_{\odot}) < 11.4$, $z < 0.4$
M2	$11.4 \leq \log_{10} (M_{\mathrm{g}}/hM_{\odot}) < 11.8$, $z < 0.4$
M3	$11.8 \leq \log_{10} (M_{\mathrm{g}}/hM_{\odot}) < 12.2$, $z < 0.4$
M4	$12.2 \leq \log_{10} (M_{\mathrm{g}}/hM_{\odot}) < 112.6$, $z < 0.4$
M5	$12.6 \leq \log_{10} (M_{\mathrm{g}}/hM_{\odot}) < 13.0$, $z < 0.4$
Blue	$u - r \leq 2.2$, $r < 20.0$
Red	$u - r > 2.2$, $r < 20.0$

Table 7.1: *Definitions of the lens galaxy samples used in Sec. 7.1.*

Name	Absolute magnitude	Name	Absolute magnitude
u_1	$-19.6 < M_u < -15.0$	i_1	$-22.0 < M_r < -17.0$
u_2	$-20.0 < M_u < -19.6$	i_2	$-22.5 < M_r < -22.0$
u_3	$-22.0 < M_u < -20.0$	i_3	$-24.0 < M_r < -22.5$
g_1	$-21.0 < M_g < -16.5$	z_1	$-22.2 < M_z < -17.0$
g_2	$-21.4 < M_g < -21.0$	z_2	$-22.6 < M_z < -22.2$
g_3	$-23.5 < M_g < -21.4$	z_3	$-24.0 < M_z < -22.6$
r_1	$-21.7 < M_r < -17.0$		For all samples:
r_2	$-22.2 < M_r < -21.7$		$14.5 < r < 17.77$
r_3	$-24.0 < M_r < -22.2$		

Table 7.2: *The lens galaxy samples used for the comparison to the SDSS (Sheldon et al. 2004).*

Sample	$\log_{10} \frac{M_{\mathrm{h}}}{h\,M_{\odot}}$	c_{h}	$\log_{10} \frac{M_{\mathrm{g}}}{h\,M_{\odot}}$	c_{g}	br	p_{c}
M1 (fit)	13.9 ± 0.2	1.4 ± 0.5	11.17 ± 0.02	6.7 ± 0.9	1.4 ± 0.2	0.88 ± 0.06
M1 (true)	13.57	4.00	11.19	7.32		0.82
M2 (fit)	13.9 ± 0.3	1.0 ± 0.5	11.55 ± 0.02	7.9 ± 0.7	1.7 ± 0.2	0.86 ± 0.09
M2 (true)	13.64	3.94	11.59	6.62		0.83
M3 (fit)	13.9 ± 0.3	1.2 ± 0.7	11.96 ± 0.02	6.7 ± 0.7	1.74 ± 0.3	0.83 ± 0.09
M3 (true)	13.71	3.86	11.98	5.98		0.86
M4 (fit)	14.0 ± 0.3	1.7 ± 0.5	12.35 ± 0.01	6.4 ± 0.2	1.86 ± 0.3	0.87 ± 0.14
M4 (true)	13.78	3.80	12.39	5.41		0.88
M5 (fit)	14.1 ± 0.3	1.0 ± 1.0	12.73 ± 0.02	6.4 ± 0.3	1.8 ± 0.4	0.89 ± 0.11
M5 (true)	13.89	3.67	12.79	4.89		0.90
Blue (fit)	14.8 ± 0.4	4.3 ± 1.5	11.49 ± 0.03	3.0 ± 0.1	2.4 ± 0.1	0.99 ± 0.02
Blue (true)	13.64	3.93	11.85	6.20		0.93
Red (fit)	14.3 ± 0.1	4.9 ± 0.8	12.35 ± 0.02	2.74 ± 0.07	2.2 ± 0.3	0.67 ± 0.17
Red (true)	13.87	3.71	12.67	5.12		0.56

Table 7.3: *Best fit parameters for the model Eq. (7.1) for the various lens galaxy samples. Also given are the "true" values inferred directly from the Millennium Simulation database.*
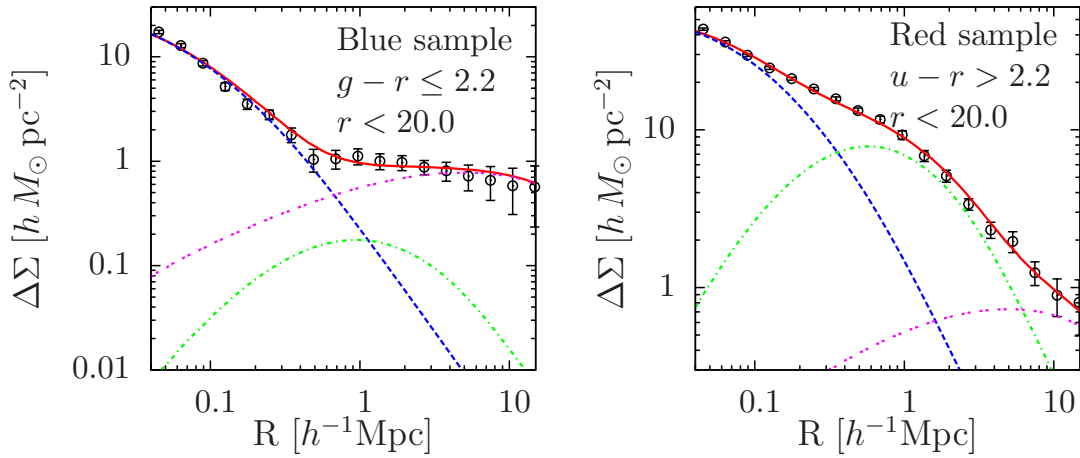
Figure 7.9: *Model fits to the GGL signals of the blue and red lens galaxy samples. The conventions for the line styles are the same as in Fig. 7.7.*

7.2 Comparison to the SDSS

So far, the most accurate measurements of galaxy-galaxy lensing have been made in the SDSS, in particular by Sheldon et al. (2004). The large area of the SDSS and the availability of spectroscopic redshifts for their lens galaxies allowed them to measure the GGL signal $\Delta\Sigma$ out to a distance of $\approx 10\,h^{-1}\,\mathrm{Mpc}$ with high signal-to-noise ratio for three different luminosity bins in all SDSS filter bands (Fig. 7.10). Since these observations comprise an ideal test ground for semi-analytic models of galaxy formation, we have simulated galaxy-galaxy lensing in mock galaxy samples that closely resemble those used in Sheldon et al. (2004). The selection criteria for the lens galaxies are summarized in Tab. 7.2. For the background population, we use randomly placed galaxies that are distributed uniformly on the sky. We use the same redshift distribution as in Sheldon et al. (2004), which was estimated by the authors using photometric redshifts. Since we are aiming for a prediction of the GGL signal from our simulations, we do neither take into account the errors on the photometric redshift estimates, nor do we assign shape noise to the background galaxies.

As discussed in Sec. 5.2.1, it is difficult to quote a single number for the small-scale resolution of the simulation. As a conservative estimate, we assume that the effects of the adaptive smoothing scheme become unimportant on scales larger than $\approx 30 - 40\,h^{-1}\,\mathrm{kpc}$.

The galaxy-galaxy lensing signals estimated from the simulations are shown in Fig. 7.11 and are compared to the SDSS data in Fig. 7.12. The differences between our results and the SDSS measurements are hardly significant – with one exception: the trend apparent in the SDSS data that galaxies with higher luminosities yield a higher $\Delta\Sigma$ is well reproduced only in the redder filter bands. In the u- and g-bands, this trend seems to reverse; the signal of the most luminous galaxies is comparable to or smaller than the signal of the intermediate samples on all scales. The reason for this behaviour can be seen in Fig. 7.14, where we show the distributions of halo masses for the different samples in the u- and r-band. While in the r-band more luminous galaxies reside in significantly more massive halos ($\log_{10}\bar{M}_{\mathrm{g}}/h\,M_\odot = 13.8$, 13.2 and 12.3 for the samples r_1,

r_2 and r_3), this is not true in the u-band, where $\log_{10} \bar{M}_{\mathrm{g}}/h\,M_\odot = 13.3$, 13.3 and 12.6 for the samples u_1, u_2 and u_3, respectively. This can explain why $\Delta\Sigma$ for the u_1-lenses is comparable to the signal from the u_2 lenses on scales below $\approx 0.5\,h^{-1}\,\mathrm{Mpc}$, where according to our model from the previous section the galaxy halo dominates the signal. Furthermore, Fig. 7.14 shows that the masses of the host halos are significantly lower in the u_1 sample ($\log_{10} \bar{M}_{\mathrm{h}}/M_\odot = 13.8$) than in the u_2-sample ($\log_{10} \bar{M}_{\mathrm{h}}/M_\odot = 14.2$) in the case of the galaxy halo being a subhalo. Since the fraction $1 - p_{\mathrm{c}}$ of such cases is approximately the same in both samples, this is the reason for the lower GGL signal of the u_1-lenses on large scales. From the information at hand, it is not possible to say with certainty which aspect of the semi-analytic model of galaxy formation is responsible for these discrepancies. However, since this problem is confined to the blue filter bands, and in addition the galaxies in the u_1-sample are significantly bluer in color than those in the r_1-sample ($u - r$=1.3 for the u_1-galaxies and $u - r$=2.2 for the r_1-galaxies), it is reasonable to suspect a problem with the recipes for star formation. Since recently formed massive stars dominate the blue light of a galaxy, the properties of the galaxies with high blue luminosities are very sensitive to the implementation of star formation in the model. In fact, Croton et al. (2006) find an excess of luminous blue galaxies when comparing the semi-analytic model to observed luminosity functions. These are galaxies that are experiencing a burst of star formation due to an ongoing merger. The authors argue that these most likely correspond to the ultraluminous infrared galaxies (ULIRGs), since they show a similar star-formation rates and abundance, with the important difference, however, that in ULIRGs most of the light from the newly formed stars in absorbed by dust and re-emitted in the infrared. The excess of luminous blue galaxies in the Croton et al. (2006) model, upon which the model by De Lucia & Blaizot (2007) is based, is therefore most likely caused by sub-optimal dust modelling. Thus, it is likely that star-forming galaxies in less massive halos have blue luminosities that are significantly too high and therefore are falsely included in the u_1 and g_1 samples, lowering the respective galaxy-galaxy lensing signals.

If we set aside the problematic u- and g-band samples, we only find differences at the level of one or two standard deviations. There is a slight trend for the SDSS measurements of $\Delta\Sigma$ from the most and least luminous lens samples to lie below the simulation predictions. Furthermore, the signal seems to be slightly steeper in the SDSS, in particular for the least luminous samples.

Two issues have to be kept in mind when comparing the simulation data to the SDSS measurements: first, the Millennium Simulation has been run with $\sigma_8 = 0.9$, a value of the power spectrum normalization which is too high compared to the current best fit value $\sigma_8 \approx 0.8$ obtained from combining CMB measurements with other cosmological probes (see Tab. 2.1). Second, the Millennium Simulation follows the evolution of the dark matter distribution only, and thus does not take into account the effects of cooling baryons and the formation of stars and galaxies. In principle, a comparison of the simulation to the SDSS data would allow us to put constraints on the amplitude of these baryonic effects. The latter, however, are partially degenerate with the effects of the high value of σ_8 in the simulation: the gravitational potential in the inner parts of a halo is dominated by the baryons condensing at the halo center, and they are therefore expected to lead to an increase of the dark matter density there (e.g. Gnedin et al. 2004, and references therein). Increasing the value of σ_8 increases the amount of matter that is associated with galaxies (e.g. Yoo et al. 2006). Therefore, at least on certain scales both effects lead to an increase of the galaxy-galaxy lensing signal. The situation is complicated further because the semi-analytic model by De Lucia & Blaizot (2007) has been tuned such that it reproduces the properties of galaxies observed in the local Universe. This approach might partially compensate for the effects of the high power spectrum normalization and the absence of

baryons in the Millennium Simulation, bringing the simulated GGL signal in better agreement with the SDSS data and making a precise comparison of simulation and observations difficult.

The dependence of $\Delta\Sigma$ on the values of Ω_m and σ_8 has been studied in Yoo et al. (2006) using the halo model and the HOD prescription. Specifically, the authors study the change of the GGL signal with σ_8 while for each value of σ_8 choosing the HOD parameters such that the angular correlation function of the model matches the SDSS observations by Zehavi et al. (2005). This situation is very similar to the Millennium Simulation, where the free parameters of the galaxy formation model have been tuned such that the properties of the simulated galaxies match the those of local galaxies in the Universe. In fact, as is shown in Springel et al. (2005), the spatial correlation function of the simulated galaxies matches the measurements from large redshift surveys, such as the 2dFGRS (Hawkins et al. 2003) or the SDSS. It is therefore plausible to estimate the effect of the high power spectrum normalization in the Millennium Simulation from the results of Yoo et al. (2006). Based on the HOD prescription, the authors present a scale-dependent fit formula of the form

$$\frac{\Delta\Sigma(R)}{\Delta\Sigma_\mathrm{fid}(R)} = \left(\frac{\Omega_\mathrm{m}}{0.3}\right)^{\alpha(R)} \left(\frac{\sigma_8}{0.8}\right)^{\beta(R)} \tag{7.17}$$

where $\Delta\Sigma_\mathrm{fid}(R)$ is the GGL signal for the fiducial cosmology with $\Omega_\mathrm{m} = 0.3$ and $\sigma_8 = 0.8$. The values given for $\beta(R)$ vary between ≈ 0.8 and 1.5 for R between 0.1 and $10\,h^{-1}\,\mathrm{Mpc}$. Even though the lens samples for which they studied these effects do not precisely match the ones considered here, we can estimate from their results that the high value of σ_8 in the simulation should lead to a GGL signal that is maximally $\approx 10 - 20\%$ higher than the signal measured in the SDSS[1].

We now investigate the effect of the lack of baryons in the Millennium Simulation. For their study of strong lensing optical depths, Hilbert et al. (2008b) added a stellar mass component at the centers of dark matter halos hosting galaxies. The authors used parametric stellar density profiles, the parameters of which were set for each galaxy according to predictions of the stellar mass and morphology of the galaxies by the semi-analytic model of De Lucia & Blaizot (2007). To balance the mass added at the halo centers, dark matter particles were removed (in an ad-hoc fashion) at the outer parts of the halos. This certainly works well in the inner parts of the halos where the stellar component dominates the mass budget. Galaxy-galaxy lensing, however, generally probes much larger scales than strong lensing, and the approach of Hilbert et al. (2008b) can therefore only be viewed as a reasonable first approximation. We use this enhanced matter distribution to produce a smaller set of 64 ray-tracing realizations to explore the importance of baryonic effects on the scales considered here. As shown in Fig. 7.13, the added stellar mass leads to a steepening of $\Delta\Sigma$ and to an increase of the GGL signal on small scales of maximally $\approx 10\%$. Overall, there is little impact on scales $\gtrsim 100\,h^{-1}\,\mathrm{kpc}$. For the least luminous lens samples, this effect improves the agreement between the simulation and the results of Sheldon et al. (2004); however, the fact that the observational errors are of a comparable size and the uncertainties of the semi-analytic model prevent us from making more quantitative statements.

In summary, we can say that our simulations and the data of Sheldon et al. (2004) agree reasonably well; the discrepancies in the blue filter are most likely due to deficiencies of the treatment of star formation and dust attenuation in the semi-analytic model. The slight over-prediction of $\Delta\Sigma$ for the samples u_3, g_3, $r_{1,3}$, $i_{1,3}$ and $z_{1,3}$ are compatible with the consequences

---

[1]Since the value of Ω_m in the simulation agrees well with the measured value quoted in Tab. 2.1, we do not expect that the increase of $\Delta\Sigma$ due to a high value σ_8 can be cancelled out this way.

expected from the high value of σ_8 in the simulation; it is not clear, however, why this is not observed for the lens galaxies with intermediate luminosities. Finally, we predict a slight steepening of the GGL signal due to the presence of baryons. Tentative evidence for this can be found for the low-luminosity lens samples, but the effect is too small to be detected with certainty on the length scales considered here.
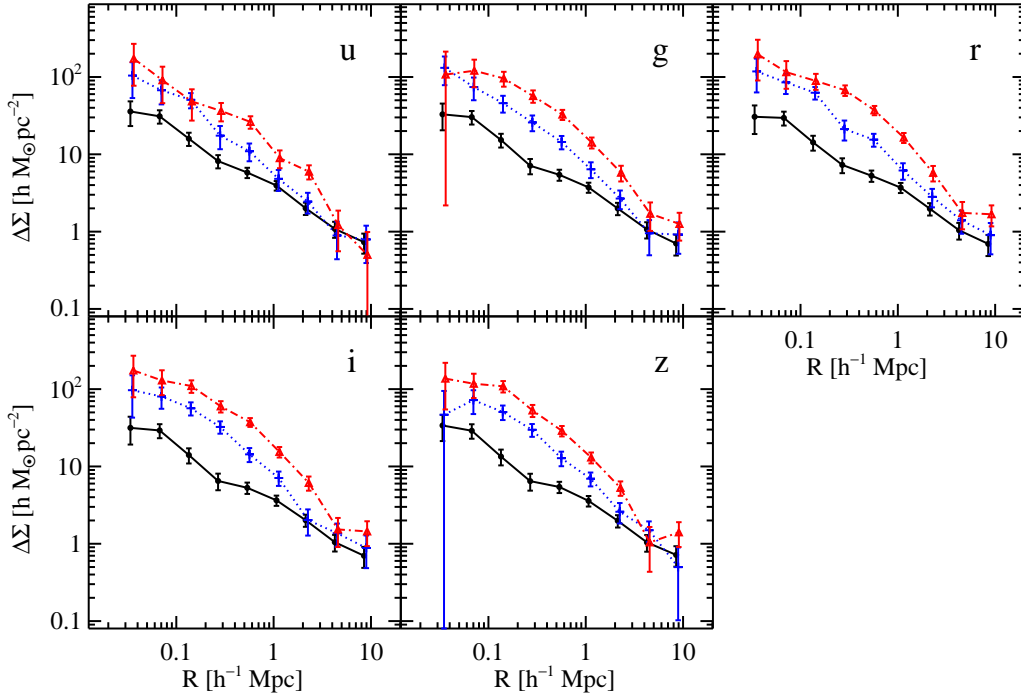
Figure 7.10: *GGL signal measured in the SDSS (figure taken from Sheldon et al. 2004). The lens galaxies were selected to be brighter than r = 17.77. They were split into three subsamples according to absolute magnitude (high, intermediate, low luminosity: dot-dashed red curves, dotted blue lines and solid black lines, respectively) in a particular filter (u, g, r, i, z).*
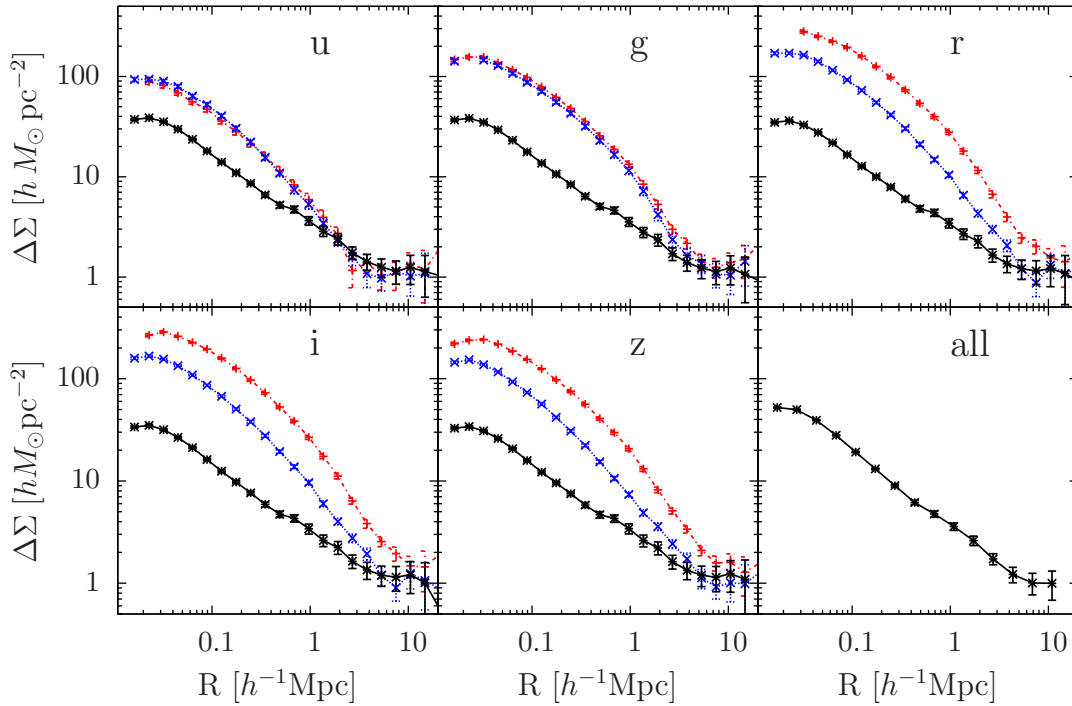
Figure 7.11: *GGL signal from the Millennium Simulation, computed for lens galaxies selected in exactly the same way as for Fig. 7.10 (high, intermediate, low luminosity: dot-dashed red curves, dotted blue lines and solid black lines, respectively). In addition to Fig. 7.10, we show in the lower right panel the signal from the joint sample containing all lens galaxies with 14.5 < r < 17.77. The error bars are estimated from the field-to-field variations.*
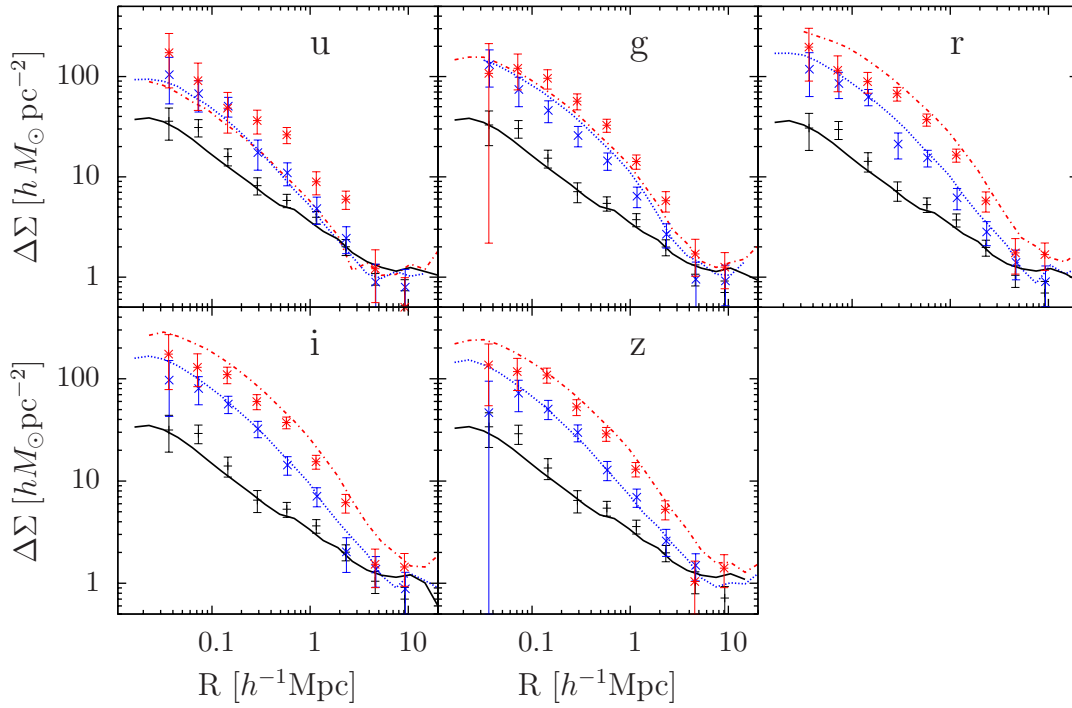
Figure 7.12: *Comparison of the GGL signals from the SDSS (Sheldon et al. 2004), shown as points with error bars, with our simulation results (lines). The data are the same as in Figs. 7.10 and 7.11 and are replotted here to facilitate the comparison.*
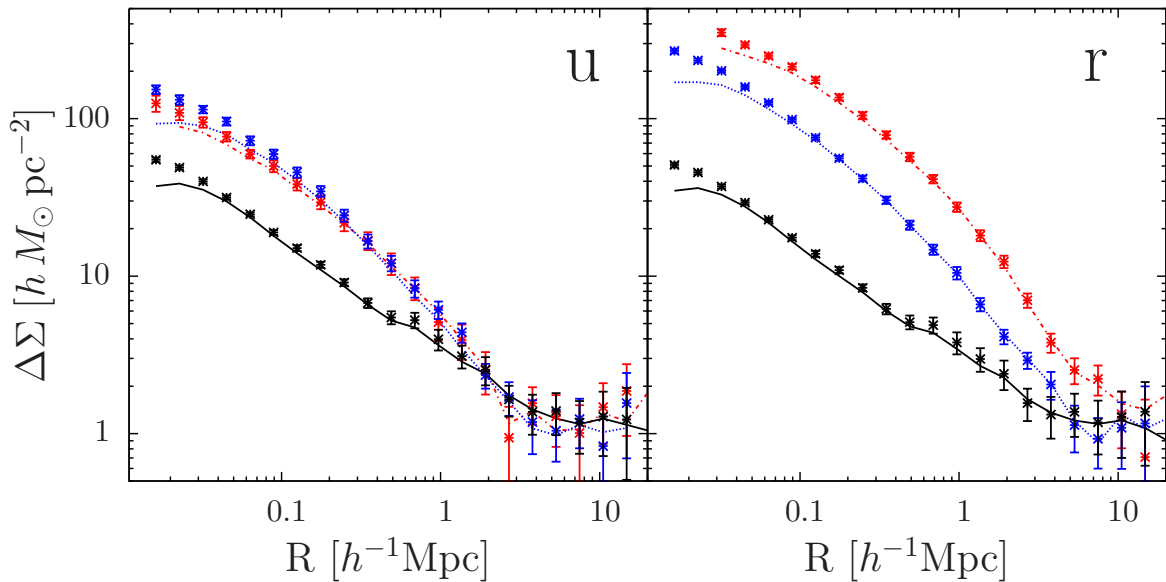
Figure 7.13: *Comparison of the GGL signals for the lens samples u_1-u_3 and r_1-r_3 computed from the dark matter simulation only (lines) and adding the stellar mass distributions of Hilbert et al. (2008b) (points with error bars).*
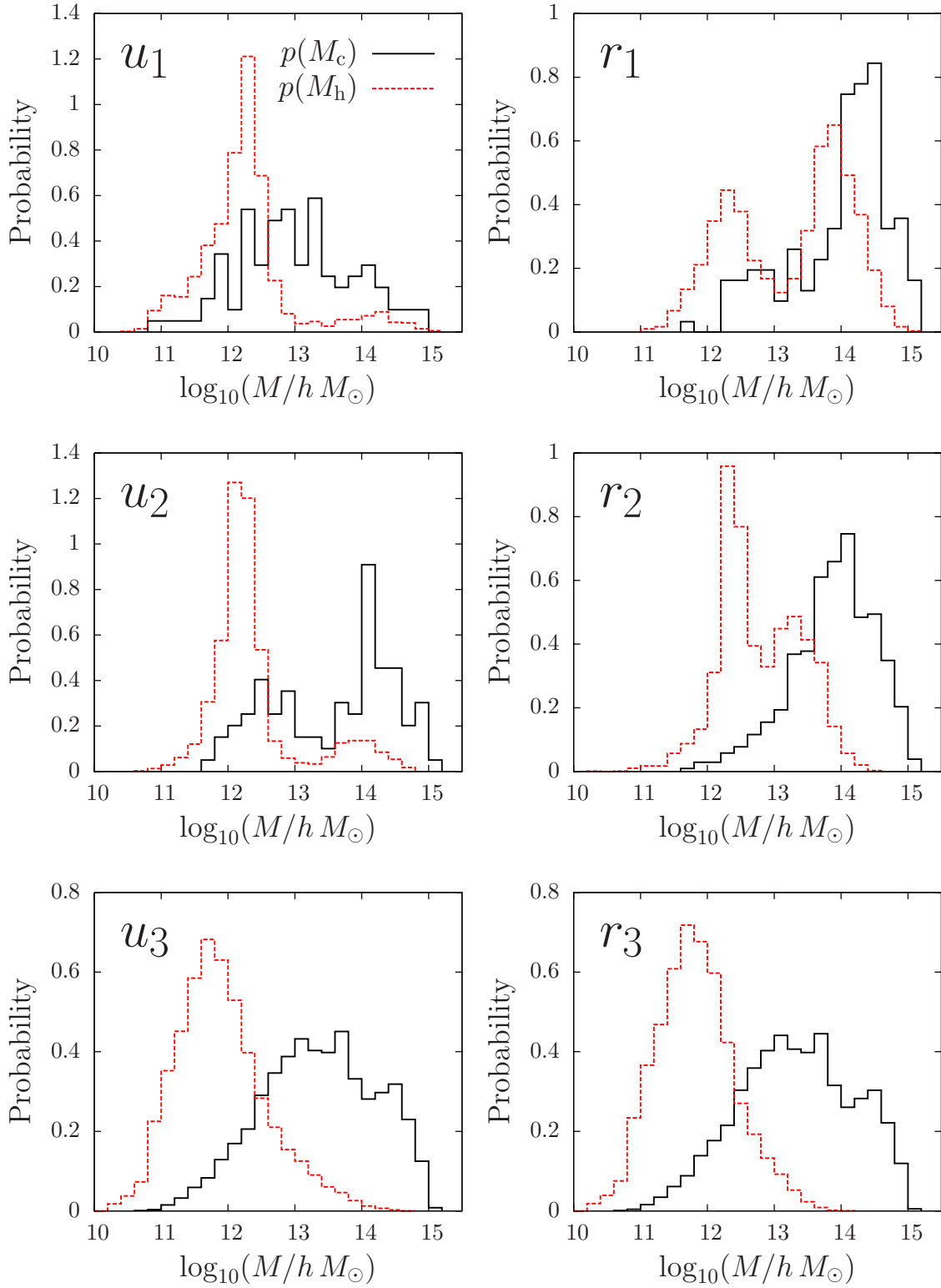
Figure 7.14: *Distributions of halo masses for the simulated SDSS lens samples in the u (left column) and r filterbands (right column). The most luminous samples are given in the top row, the least luminous in the bottom row. In each panel, we show the distributions of the masses of the lens galaxy halos (p(M_g), dashed red histograms) and the masses of the host halos if the galaxy halo is a subhalo (p(M_h), solid black histograms).*

8 Bias and correlation factors from weak lensing

The formation of galaxies is closely linked to the underlying dark matter distribution; baryons condense at the centers of dark matter halos, and therefore it can be expected that the galaxy distribution in some way traces the overall matter distribution in the Universe. However, it cannot be expected that this relation is simple. The local density is certainly not the only parameter controlling galaxy formation: for example, the ability of the pre-galactic gas to cool depends on its metallicity, and the cooling is most likely regulated by feedback processes involving supernovae and quasar activity. Furthermore, galaxies with different properties are found to be distributed in different ways; red, early type galaxies are predominantly found in dense environments, whereas blue, late-type galaxies are mostly field galaxies (e.g. Goto et al. 2003).

Therefore, galaxies are biased tracers of the distribution of the dark matter. Not only does this *galaxy bias* encode information about the environmental dependence of the processes involved in the formation of galaxies, but a detailed understanding of this effect is vital for the precise determination of cosmological parameters from the distribution of luminous matter, such as galaxy redshift surveys (e.g. Percival et al. 2007).

Since weak gravitational lensing is sensitive to the overall matter distribution, the galaxy bias can be studied by combining it with observations of the luminous matter. In this Chapter, we investigate the possibilities and limitations of this approach, using a method that was proposed in van Waerbeke (1998) and Schneider (1998) and was extended and successfully applied to the Red-Sequence Cluster Survey (RCS) by Hoekstra et al. (2001, 2002) and to the GaBoDS survey by Simon et al. (2007).

8.1 The relation of galaxies and dark matter

Several suggestions have been made in order to quantify the relation between galaxies and dark matter. In this section, we only describe the parametrization relevant to this work, the *linear stochastic bias* following Dekel & Lahav (1999) and Tegmark & Peebles (1998), and refer the reader to Simon (2005) for a thorough review of the topic.

We denote by δ_R the three-dimensional matter density contrast smoothed with a filter function W_R with filter scale R, and correspondingly by $\delta_{\mathrm{g},R}$ the smoothed galaxy density fluctuation field. Both are homogeneous and isotropic random fields (see Sec. 2.2.3), which we treat as components of the vector $\boldsymbol{d} = (\delta_R, \delta_{\mathrm{g},R})^{\mathrm{t}}$. The probability distribution of $\boldsymbol{d}$ contains all information about the relation of $\tilde{\delta}_R$ and $\tilde{\delta}_{\mathrm{g},R}$ for a fixed smoothing scale R. By definition of the density contrasts, the first non-vanishing moment of $p(\boldsymbol{d})$ is the covariance matrix, which we write as

$$\langle \boldsymbol{d}\boldsymbol{d}^{\mathrm{t}} \rangle = \sigma_R^2 \begin{pmatrix} 1 & b(R)\,r(R) \\ b(R)\,r(R) & b^2(R) \end{pmatrix} , \tag{8.1}$$

where σ_R^2 is the variance of δ_R. In Eq. (8.1), we have defined the two *linear bias parameters* $b(R)$ (the *bias factor*) and $r(R)$ (the *correlation factor*). If δ and δ_{g} were Gaussian random fields,

these two quantities would be sufficient to fully describe the relation between galaxies and dark matter. While this is approximately true for large smoothing scales, higher moments of $p(\boldsymbol{d})$ become important if $|\delta| \gtrsim 1$.

For what follows, it will be convenient to work in Fourier space. The second moment of $p_{\boldsymbol{k}}(\tilde{\boldsymbol{d}})$, the joint probability distribution function of $\tilde{\boldsymbol{d}} = (\tilde{\delta}, \tilde{\delta}_{\mathrm{g}})^{\mathrm{t}}$, can be written in analogy to Eq. (8.1) as

$$\left\langle \tilde{\boldsymbol{d}}(\boldsymbol{k})\, \tilde{\boldsymbol{d}}^{\mathrm{t}}(\boldsymbol{k}') \right\rangle = (2\pi)^3\, \delta_{\mathrm{D}}\left(\boldsymbol{k} - \boldsymbol{k}'\right)\, P_\delta(|\boldsymbol{k}|) \begin{pmatrix} 1 & b(k)\, r(k) \\ b(k)\, r(k) & b^2(k) \end{pmatrix} , \tag{8.2}$$

so that

$$b(k) = \sqrt{\frac{P_{\mathrm{g}}(k)}{P_\delta(k)}} , \tag{8.3}$$

$$r(k) = \frac{P_{\delta,\mathrm{g}}(k)}{\sqrt{P_{\mathrm{g}}(k)\, P_\delta(k)}} . \tag{8.4}$$

Thus, $b(k)$ parametrizes the scale-dependent ratio of the clustering strength of galaxies and dark matter, whereas the correlation factor $r(k)$ is a measure of the stochasticity in the relation between δ and δ_{g}. The latter takes into account that the process of galaxy formation and evolution does not only depend on the local dark matter density, but also on many other hidden parameters. For example, the efficiency with which gas can cool and form stars depends on its chemical composition; furthermore, the properties of present day galaxies are the results of the individual mass accretion histories, which in turn depend on the environment of the galaxy halo, etc. For these reasons it cannot be expected that there exists a deterministic function $\delta_{\mathrm{g}}(\delta)$. However, as shown in Dekel & Lahav (1999), r is also sensitive to non-linearities in the relation $\delta_{\mathrm{g}}(\delta)$, which become important in particular on small scales where the fields can no longer be approximated as being Gaussian.

As has been noted in Simon et al. (2007), the relation between the bias parameters in real and Fourier space can be written down once a filter function W_R has been specified:

$$b^2(R) = \frac{\int \mathrm{d}k\, k^2\, b^2(k)\, P_\delta(k)\, |\tilde{W}_R(k)|^2}{\int \mathrm{d}k\, k^2\, P_\delta(k)\, |\tilde{W}_R(k)|^2} , \tag{8.5}$$

$$r(R) = \frac{1}{b(R)} \frac{\int \mathrm{d}k\, k^2\, b(k)\, r(k)\, P_\delta(k)\, |\tilde{W}_R(k)|^2}{\int \mathrm{d}k\, k^2\, P_\delta(k)\, |\tilde{W}_R(k)|^2} , \tag{8.6}$$

where $\tilde{W}_R(k)$ is the Fourier transform of W_R.

8.2 Aperture statistics

To obtain the linear stochastic bias parameters b and r, the power spectra P_{g}, P_δ and $P_{\delta\mathrm{g}}$ or related quantities have to be measured. The strategy that we will follow in this Chapter is as follows: we divide the galaxy catalogue into a fore- and a background sample. The cosmic shear signal of the background galaxies can be used to infer the projected matter power spectrum P_κ. The galaxy-galaxy lensing signal of the foreground sample yields the cross-correlation of the lenses with the dark matter distribution, whereas the angular clustering power spectrum P_ω can be obtained directly from the positions of the foreground galaxies.

However, it is difficult to measure the projected power spectra directly. As pointed out before, the aperture mass dispersion (Eq. 3.91) provides a very convenient measure of P_κ, which can be computed from the shear correlation functions and for this reason are much easier to estimate. Therefore, we now define aperture statistics analogous to M_{ap}^2 that are measures of the angular clustering of the lens galaxies and of their cross-correlation to the dark matter distribution.

Using the notation of Sec. 3.4, the fractional number density of galaxies is defined as

$$\kappa_{\mathrm{g}}(\boldsymbol{\theta}) = \frac{n_{\mathrm{g}}(\boldsymbol{\theta}) - \bar{n}}{\bar{n}} = \int \mathrm{d}w \, p_{\mathrm{f}}(w) \, \delta_{\mathrm{g}}[f_K(w)\boldsymbol{\theta}, w] \; , \tag{8.7}$$

where $n_{\mathrm{g}}(\boldsymbol{\theta})$ is the number density of lens galaxies on the sky, $\bar{n}$ is the projected mean number density and $p_{\mathrm{f}}(w)$ is the distribution of the lenses in comoving distance. Note that we are working in the Born approximation throughout.

With this, the aperture number count statistic can be defined in analogy to M_{ap}:

$$\mathcal{N}(\boldsymbol{\vartheta}) = \int \mathrm{d}^2\vartheta' \, U_\theta(|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'|) \, \kappa_{\mathrm{g}}(\boldsymbol{\vartheta}') \; , \tag{8.8}$$

where the aperture defined by U has a radius of θ and is centered on $\boldsymbol{\vartheta}$.

With this definition, we can now define the following aperture dispersions:

$$\langle \mathcal{N}^2(\theta) \rangle = \int \mathrm{d}^2\vartheta \, U_\theta(|\boldsymbol{\vartheta}|) \int \mathrm{d}^2\vartheta' \, U_\theta(|\boldsymbol{\vartheta}'|) \, \langle \kappa_{\mathrm{g}}(\boldsymbol{\vartheta})\kappa_{\mathrm{g}}(\boldsymbol{\vartheta}') \rangle \; , \tag{8.9}$$

$$\langle M_{\mathrm{ap}}^2(\theta) \rangle = \int \mathrm{d}^2\vartheta \, U_\theta(|\boldsymbol{\vartheta}|) \int \mathrm{d}^2\vartheta' \, U_\theta(|\boldsymbol{\vartheta}'|) \, \langle \kappa(\boldsymbol{\vartheta})\kappa(\boldsymbol{\vartheta}') \rangle \; , \tag{8.10}$$

$$\langle \mathcal{N}(\theta) M_{\mathrm{ap}}(\theta) \rangle = \int \mathrm{d}^2\vartheta \, U_\theta(|\boldsymbol{\vartheta}|) \int \mathrm{d}^2\vartheta' \, U_\theta(|\boldsymbol{\vartheta}'|) \, \langle \kappa_{\mathrm{g}}(\boldsymbol{\vartheta})\kappa(\boldsymbol{\vartheta}') \rangle . \tag{8.11}$$

The correlation functions $\langle \kappa\,\kappa \rangle$ in the integrands are the Fourier transforms of the angular clustering power spectrum P_ω, the convergence power spectrum P_κ and the cross power spectrum $P_{\kappa\mathrm{g}}$, respectively. Since our goal is to relate these statistics to the three-dimensional bias parameters, we express these projected power spectra by their three-dimensional counterparts using Limber's equation (Limber 1953): let

$$g_i(\boldsymbol{\theta}) = \int \mathrm{d}w \, q_i(w) \, \delta_i[f_K(w)\boldsymbol{\theta}, w] \tag{8.12}$$

be projections of the three-dimensional density contrasts δ_i with weight functions q_i, $i = 1, 2$. Then, Limber's equation states that the (cross-) power spectra of g_1 and g_2 are given by

$$P_{\mathrm{2D},ij}(\ell) = \int \mathrm{d}w \, \frac{q_i(w) \, q_j(w)}{f_K^2(w)} \, P_{ij}\left(\frac{\ell}{f_K(w)}, w\right) \; . \tag{8.13}$$

Reading off the appropriate weight functions for κ and κ_{g} from Eqs. (3.67) and (8.7), the projected power spectra are given by

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

$$P_\omega(\ell) = \int \mathrm{d}w \, \frac{p_{\mathrm{f}}^2(w)}{f_K^2(w)} \, b^2\left(\frac{\ell}{f_K(w)}, w\right) P_\delta\left(\frac{\ell}{f_K(w)}, w\right) ,$$

$$P_{\kappa\mathrm{g}}(\ell) = \frac{3H_0^2 \Omega_{\mathrm{m}}}{2c^2} \int \mathrm{d}w \, \frac{g(w)p_{\mathrm{f}}(w)}{a(w)f_K(w)} \, b\left(\frac{\ell}{f_K(w)}, w\right) r\left(\frac{\ell}{f_K(w)}, w\right) P_\delta\left(\frac{\ell}{f_K(w)}, w\right) .$$

In the latter two equations, Eqs. (8.3) and (8.4) were used to express P_g and $P_{\delta\mathrm{g}}$ in terms P_δ and the linear bias parameters.

In analogy to Sec. 3.3.4, the aperture statistics defined in Eqs. (8.9)-(8.11) can now be related to the three-dimensional power spectra and the bias and correlation factors:

$$
\begin{aligned}
\left\langle \mathcal{N}^2(\theta) \right\rangle &= \frac{1}{2\pi} \int \mathrm{d}\ell\, \ell\, P_\omega(\ell)\, W_\mathrm{ap}^2(\theta\ell) & (8.14) \\
&= \frac{1}{2\pi} \int \mathrm{d}k\, k \int \mathrm{d}w\, p_\mathrm{f}^2(w)\, b^2\,(k;\, w)\, P_\delta\,(k;\, w)\, W_\mathrm{ap}^2\,[f_K(w)\theta\, k] \ , \\
\left\langle M_\mathrm{ap}^2(\theta) \right\rangle &= \frac{1}{2\pi} \int \mathrm{d}\ell\, \ell\, P_\kappa(\ell)\, W_\mathrm{ap}^2(\theta\ell) & (8.15) \\
&= \frac{9H_0^4\Omega_\mathrm{m}^2}{4c^4} \frac{1}{2\pi} \int \mathrm{d}k\, k \int \mathrm{d}w\, f_K^2(w)\frac{g^2(w)}{a^2(w)}\, P_\delta\,(k;\, w)\, W_\mathrm{ap}^2\,[f_K(w)\theta\, k] \ , \\
\left\langle \mathcal{N}(\theta) M_\mathrm{ap}(\theta) \right\rangle &= \frac{1}{2\pi} \int \mathrm{d}\ell\, \ell\, P_{\kappa\mathrm{g}}(\ell)\, W_\mathrm{ap}^2(\theta\ell) & (8.16) \\
&= \frac{3H_0^2\Omega_\mathrm{m}}{2c^2} \frac{1}{2\pi} \int \mathrm{d}k\, k \int \mathrm{d}w\, f_K(w)\frac{g(w)p_\mathrm{f}(w)}{a(w)}\, b\,(k;\, w)\, r\,(k;\, w) \\
&\quad \times P_\delta\,(k;\, w)\, W_\mathrm{ap}^2\,[f_K(w)\theta\, k] \ .
\end{aligned}
$$

For the polynomial weight function U_θ (Eq. 3.94) for the aperture statistic, the filter function W_ap^2 is given by $W_\mathrm{ap}^2(\eta) = 576\, \mathrm{J}_4^2(\eta)/\eta^4$.

Eqs. (8.14)-(8.16) are useful to make theoretical predictions of the aperture statistics. Since the power spectra in these equations are hard to measure in practice, it is desireable to express the aperture statistics in terms of more easily observable quantities. A relation of $\left\langle M_\mathrm{ap}^2 \right\rangle$ to the shear correlation functions has already been given in Eq. (3.93). Similarly, $\left\langle \mathcal{N}^2 \right\rangle$ can be expressed in terms of the angular correlation function $\omega(\vartheta)$ of the lens galaxies:

$$
\left\langle \mathcal{N}^2(\theta) \right\rangle = \int_0^{2\theta} \frac{\mathrm{d}\vartheta\, \vartheta}{\theta^2}\, \omega(\vartheta)\, T_+\left(\frac{\vartheta}{\theta}\right) , \tag{8.17}
$$

where T_+ is given by Eq. (3.96) or Eq. (3.101) for the polynomial or exponential weight functions U_θ, respectively. We compute $\omega(\vartheta)$ using the Landy-Szalay estimator (Landy & Szalay 1993)

$$
\omega(\vartheta) = \frac{DD(\vartheta) - 2DR(\vartheta) + RR(\vartheta)}{RR(\vartheta)} \ . \tag{8.18}
$$

Here, $DD(\vartheta)$ is the number of pairs of galaxies in the data set that fall into the separation bin centered on ϑ, and RR is the number of pairs in a mock catalog of randomly placed galaxies with a uniform distribution, which has the same geometry as the data catalog. Finally, $DR(\vartheta)$ is the number of pairs that can be formed using one galaxy from the data set and one from the mock catalog.

In an analogous way, $\left\langle \mathcal{N} M_\mathrm{ap} \right\rangle$ can be obtained from the galaxy-galaxy lensing signal (see Sec. 3.4):

$$
\left\langle \mathcal{N}(\theta) M_\mathrm{ap}(\theta) \right\rangle = \int_0^{2\theta} \frac{\mathrm{d}\vartheta\, \vartheta}{\theta^2}\, \left\langle \gamma_\mathrm{t}(\theta) \right\rangle T_2\left(\frac{\vartheta}{\theta}\right) , \tag{8.19}
$$

where, for the polynomial filter function, T_2 is given by

$$
T_2(x) = 567 \int_0^\infty \frac{\mathrm{d}t}{t^3}\, \mathrm{J}_2(xt)\, [\mathrm{J}_4(t)]^2 \ . \tag{8.20}
$$

8.3 Recovering bias and correlation factor

Since we wish to construct functions of the aperture statistics that are as close as possible to the linear bias parameters, b and r have to be extracted from the integrals in Eqns. (8.14)–(8.16). Clearly, this is only possible for averages of b and r. We begin by defining the k-averages of b^2 and br:

$$\overline{b^2}\left[f_K(w)\theta; w\right] = \int_0^\infty dk\, W_k(k, \theta, w)\, b^2(k; w)\;, \tag{8.21}$$

$$\overline{br}\left[f_K(w)\theta; w\right] = \int_0^\infty dk\, W_k(k, \theta, w)\, b(k; w)\, r(k; w)\;, \tag{8.22}$$

where

$$W_k(k, \theta, w) = \frac{k\, P_\delta(k; w)\, W_{\mathrm{ap}}^2\left[f_K(w)\theta\, k\right]}{\int_0^\infty dk\, k\, P_\delta(k; w)\, W_{\mathrm{ap}}^2\left[f_K(w)\theta\, k\right]}\;. \tag{8.23}$$

These averages account for the fact that since the aperture statistics only measure a smoothed version of the power spectrum, only a smoothed version of the bias parameters can be recovered. The relevant aperture dispersions can then be written as

$$\left\langle \mathcal{N}^2(\theta)\right\rangle = \frac{1}{2\pi}\int dw\, \frac{p_{\mathrm{f}}^2(w)}{f_K^2(w)}\, \overline{b^2}\left[f_K(w)\theta; w\right]\, \mathcal{P}(w, \theta) \qquad \text{and} \tag{8.24}$$

$$\left\langle \mathcal{N}(\theta) M_{\mathrm{ap}}(\theta)\right\rangle = \frac{3H_0^2\Omega_{\mathrm{m}}}{2c^2}\frac{1}{2\pi}\int dw\, \frac{g(w)p_{\mathrm{f}}(w)}{a(w)f_K(w)}\, \overline{br}\left[f_K(w)\theta; w\right]\, \mathcal{P}(w, \theta)\;, \tag{8.25}$$

where we have introduced

$$\mathcal{P}(w, \theta) \equiv \int dk\, k\, f_K^2(w)\, P_\delta(k, w)\, W_{\mathrm{ap}}[\theta k f_K(w)]\;. \tag{8.26}$$

For notational convenience, we define the following auxiliary functions:

$$h_{\kappa\kappa}(w, \theta) \equiv \frac{9H_0^4\Omega_{\mathrm{m}}^2}{4c^4}\frac{g^2(w)}{a^2(w)}\mathcal{P}(w, \theta)\;, \tag{8.27}$$

$$h_{\mathrm{gg}}(w, \theta) \equiv \frac{p_{\mathrm{f}}^2(w)}{f_K^2(w)}\mathcal{P}(w, \theta) \qquad \text{and} \tag{8.28}$$

$$h_{\kappa\mathrm{g}}(w, \theta) \equiv \frac{3H_0^2\Omega_{\mathrm{m}}}{2c^2}\frac{g(w)p_{\mathrm{f}}(w)}{f_K(w)a(w)}\mathcal{P}(w, \theta)\;. \tag{8.29}$$

Since the angular clustering and the GGL signal of the lens galaxies are averages over the lens redshift distribution, the resulting bias parameters will be redshift averages, too. Thus, we define

$$\left[\overline{b^2}\right]_{\mathrm{gg}}(\theta) = \frac{\int dw\, \overline{b^2}\left[f_K(w)\theta; w\right] h_{\mathrm{gg}}(w, \theta)}{\int dw\, h_{\mathrm{gg}}(w, \theta)} \tag{8.30}$$

$$\left[\overline{br}\right]_{\kappa\mathrm{g}}(\theta) = \frac{\int dw\, \overline{br}\left[f_K(w)\theta; w\right] h_{\kappa\mathrm{g}}(w, \theta)}{\int dw\, h_{\kappa\mathrm{g}}(w, \theta)}\;. \tag{8.31}$$

This finally allows us to pull the averages $\left[\overline{b^2}\right]_{\mathrm{gg}}$ and $\left[\overline{br}\right]_{\kappa\mathrm{g}}$ out of the integrals over k and w, so that we can write

$$\left\langle \mathcal{N}^2(\theta) \right\rangle = \frac{1}{2\pi} \left[\overline{b^2}\right]_{\mathrm{gg}}(\theta) \int \mathrm{d}w\, h_{\mathrm{gg}}(w,\theta)\,, \tag{8.32}$$

$$\left\langle \mathcal{N}(\theta) M_{\mathrm{ap}}(\theta) \right\rangle = \frac{1}{2\pi} \left[\overline{br}\right]_{\kappa\mathrm{g}}(\theta) \int \mathrm{d}w\, h_{\kappa\mathrm{g}}(w,\theta) \qquad \text{and} \tag{8.33}$$

$$\left\langle M_{\mathrm{ap}}^2(\theta) \right\rangle = \frac{1}{2\pi} \int \mathrm{d}w\, h_{\kappa\kappa}(w,\theta)\,. \tag{8.34}$$

By rearranging Eqns. (8.32)–(8.34), we can define the projected bias parameters

$$b_{\mathrm{2D}}(\theta) \equiv \sqrt{\left[\overline{b^2}\right]_{\mathrm{gg}}(\theta)} = \sqrt{\frac{\left\langle \mathcal{N}^2(\theta) \right\rangle}{\left\langle M_{\mathrm{ap}}^2(\theta) \right\rangle}}\, f_b(\theta)\,, \tag{8.35}$$

$$r_{\mathrm{2D}}(\theta) \equiv \frac{\left[\overline{br}\right]_{\kappa\mathrm{g}}(\theta)}{\sqrt{\left[\overline{b^2}\right]_{\mathrm{gg}}(\theta)}} = \frac{\left\langle \mathcal{N}(\theta) M_{\mathrm{ap}}(\theta) \right\rangle}{\sqrt{\left\langle \mathcal{N}^2(\theta) \right\rangle \left\langle M_{\mathrm{ap}}^2(\theta) \right\rangle}}\, f_r(\theta)\,, \tag{8.36}$$

where

$$f_b(\theta) = \frac{\int \mathrm{d}w\, h_{\kappa\kappa}(w,\theta)}{\int \mathrm{d}w\, h_{\mathrm{gg}}(w,\theta)}\,, \tag{8.37}$$

$$f_r(\theta) = \frac{\sqrt{\int \mathrm{d}w\, h_{\kappa\kappa}(w,\theta)}\sqrt{\int \mathrm{d}w\, h_{\mathrm{gg}}(w,\theta)}}{\int \mathrm{d}w\, h_{\kappa\mathrm{g}}(w,\theta)}\,. \tag{8.38}$$

The aperture dispersions serve as proxies for the power spectra, and thus the functions b_{2D} and r_{2D} can be considered to be the two-dimensional analogs of the definitions of the three-dimensional bias parameters (see Eqs. 8.3 and 8.4). Note that the integrals over $h_{\kappa\kappa}$, h_{gg} and $h_{\kappa\mathrm{g}}$ in Eqs. (8.37) and (8.38) are just the aperture dispersions $\left\langle M_{\mathrm{ap}}^2(\theta) \right\rangle$, $\left\langle \mathcal{N}^2(\theta) \right\rangle$ and $\left\langle \mathcal{N}(\theta) M_{\mathrm{ap}}(\theta) \right\rangle$ for the case of unbiased ($b = r = 1$) galaxies. Since f_b and f_r can be computed theoretically from the dark matter power spectrum, b_{2D} and r_{2D} can be obtained directly from measurements of the angular clustering and the GGL signal of the lenses and the cosmic shear signal of galaxies in the background. As in Chapter 5, we will base our computations of f_b and f_r on the three-dimensional power spectrum measured in the Millennium Simulation instead of resorting to fit formulae.

8.4 Predicting the two-dimensional bias parameters

The two-dimensional bias parameters given in Eqs. (8.35) and (8.36) are rather complicated functions of b and r, in particular because in general $\overline{br} \neq \overline{b}\,\overline{r}$, $\sqrt{\left[\overline{b^2}\right]_{\mathrm{gg}}} \neq \left[\overline{b}\right]_{\mathrm{gg}}$ and $[\cdot]_{\mathrm{gg}} \neq [\cdot]_{\kappa\mathrm{g}}$. However, Eqs. (8.35) and (8.36) can in principle be used to predict the projected bias parameters for any desired lens galaxy sample even if no ray-tracing simulations are available, provided the redshift distributions of lens and source galaxies and the three-dimensional bias factors $b(k)$ and $r(k)$ are known. The latter can be measured directly from the N-body simulation and the semi-analytic galaxy formation model using Eqs. (8.3) and (8.4). In this section, we investigate

this approach in more detail. We compare the predictions obtained in this way to the results obtained from the ray-tracing simulations using the weak lensing method in Sec. 8.5.

To estimate the three-dimensional b and r, we first estimate the three-dimensional power spectra P_δ, P_g and $P_{\delta g}$ using the "*chaining-the-power*" technique proposed by Jenkins et al. (1998) and Smith et al. (2003), which we have also discussed in detail in Hartlap (2005). We then use Eqs. (8.3) and (8.4) to obtain estimates of $b(k)$ and $r(k)$. This can only be done for a limited range in k. On large scales, the limitation arises from the finite size of the simulation box; the bias parameters cannot be estimated for $k \lesssim 2\pi/L$. The limiting factor on small scales is the galaxy power spectrum. Since the number of galaxies is generally much smaller than the number of dark matter particles, shot noise becomes important first for P_g. The scale k_{shot} where shot noise begins to dominate the power spectrum can be estimated by solving $P_g(k_{shot}) = 1/n_g$, where n_g is the number density of galaxies in the simulation. Even if the shot noise is subtracted from the power spectrum, it is generally not possible to obtain a meaningful estimate of the power spectrum significantly beyond this scale. In particular, it seems that the shot noise correction based on the assumption that the galaxy distribution results from the Poisson sampling of an underlying continuous density field is slightly too large, resulting in a negative power spectrum (see also, e.g., Desjacques et al. 2008).

In Fig. 8.1, we show as an example the bias parameters estimated using Eqs. (8.3) and (8.4) for three different redshifts for the blue and the red galaxy sample. While the blue sample is anti-biased ($b < 1$) with respect to the dark matter, the bias factor for the red sample is generally larger than unity. This is consistent with the fact that red galaxies are predominantly found in groups and clusters, whereas the blue galaxies are mostly field galaxies. Note that the increase of both bias and correlation factor on small scales is a real feature of the simulation; it can be explained in terms of the halo model (see e.g. Seljak 2000). Furthermore, the fact that the correlation factor exceeds unity in some cases is a consequence the shot-noise correction applied to the galaxy power spectrum. This can be seen as follows: the shot noise contribution is negligible for the estimated dark matter power spectrum $\hat{P}_\delta$ (due to the large number of dark matter particles in the Millennium Simulation) and is completely absent in the estimated cross power spectrum $\hat{P}_{\kappa g}$. Our estimate of the correlation factor can therefore be written as

$$\hat{r}(k) = \frac{\hat{P}_{\kappa g}(k)}{\sqrt{\left[\hat{P}_g(k) - 1/n_g\right] \hat{P}_\delta(k)}} = \frac{\hat{P}_{\kappa g}(k)}{\sqrt{\hat{P}_g(k)\,\hat{P}_\delta(k) - \hat{P}_g(k)/n_g}} \ . \tag{8.39}$$

The term $\hat{P}_g(k)/n_g$ is most important for large k, because $\hat{P}_g\,\hat{P}_\delta$ declines more quickly than $\hat{P}_g$ as k increases, and can therefore cause the correlation factor to become larger than unity on these scales.

Fig. 8.1 also illustrates the limits imposed by the finite number of galaxies: the value of k to which we can estimate the bias parameters decreases with increasing redshift. The reason for this is that due to the flux limit of $r < 20$ there are fewer galaxies at higher redshifts, and therefore the shot noise comes to dominate on larger scales. Furthermore, since the blue galaxies show very little clustering, shot noise is relatively more important than for the red galaxies. This explains that the bias parameter estimates are much better for the red sample, even though there is approximately the same number of galaxies in each sample.

Given the three-dimensional bias parameters, we can now in principle use the first equalities in Eqs. (8.35) and (8.36) to predict b_{2D} and r_{2D}. However, as discussed above, the galaxy power spectrum cannot always be estimated over the range in k that would be necessary to carry out

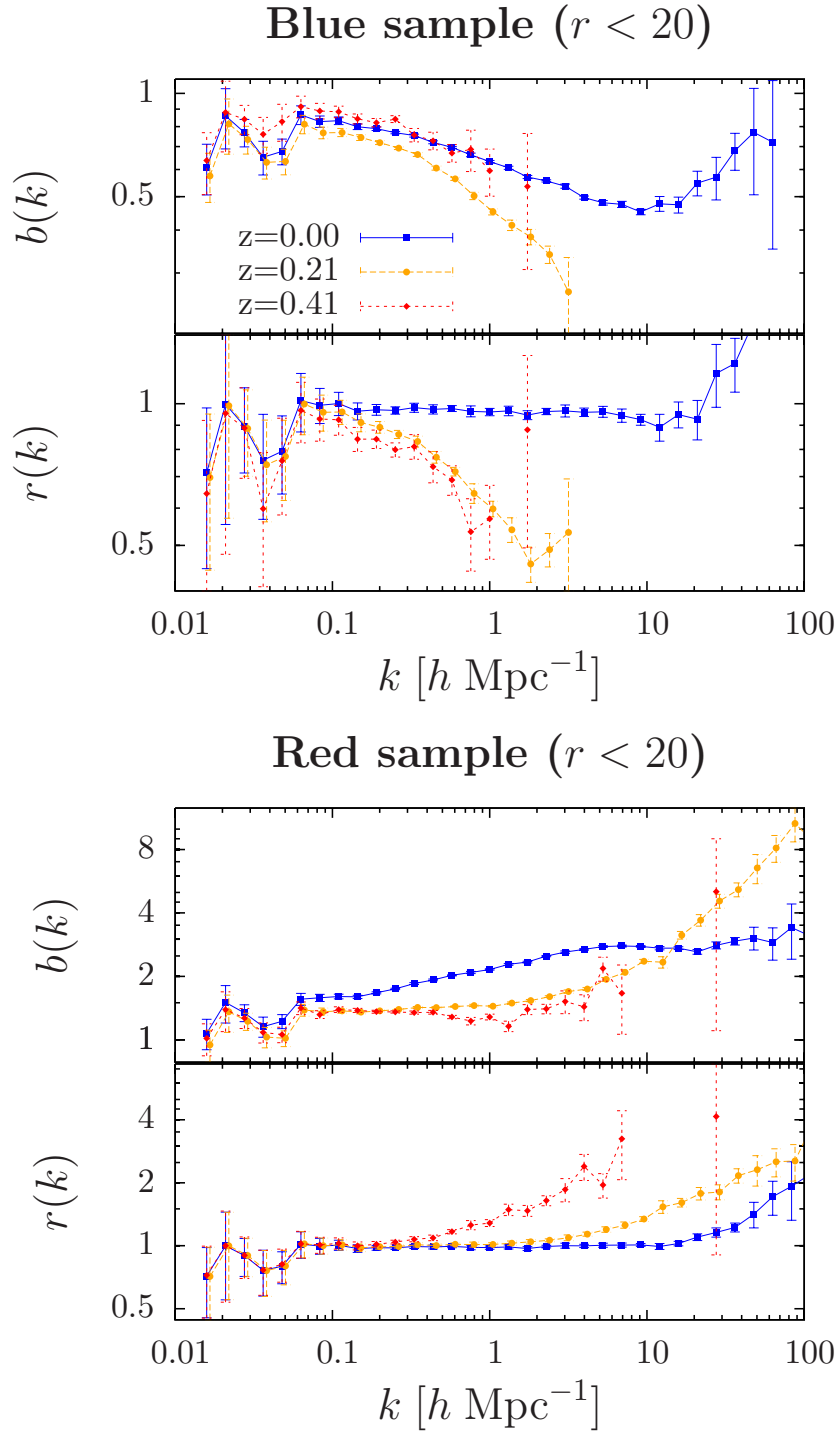## Blue sample ($r < 20$)

Red sample ($r < 20$)

Figure 8.1: *Three-dimensional bias parameters for the red and blue samples (with $r < 20$). For better visibility, the data points have been connected by lines. We only plot the bias parameters up to the scale where shot-noise begins to dominated the galaxy power spectrum. The error bars were computed by error propagation from the errors on the power spectra, which in turn reflect the scatter within the individual k-bins.*
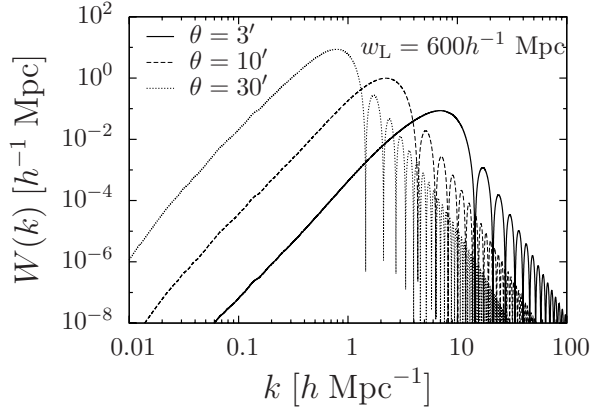
Figure 8.2: *The filter function $W_k(k)$ (Eq. 8.23) for three different aperture radii.*
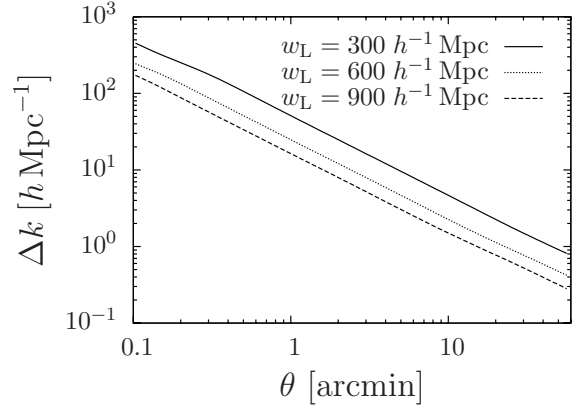
Figure 8.3: *Width of the filter $W_k(k)$ as function of the aperture radius, for three different lens redshifts. Here, the width Δk is defined as the interval that contains 90% of the total weight.*

the averaging operations over the function W_k in Eqs. (8.21) and (8.22). In Fig. 8.2, we plot the weight function W_k for three different angular scales and lens galaxies at $w_L = 600\,h^{-1}$ Mpc. We define the width Δk of W_k as size of the symmetric interval centred on the maximum of W_k which contains 90% of the weight. Note that the apparent scaling behaviour $\Delta k \propto 1/\theta$ is only approximate due to the presence of the power spectrum in Eq. (8.23).

As can be seen from Fig. 8.3, the filter function is wider for smaller aperture radii and lower lens redshifts. Since for smaller aperture radii the filter function will eventually pick up contributions from the shot-noise dominated regime at large k (see Fig. 8.2), there is a fundamental limit to the accuracy of predictions for realistic lens galaxy samples on small scales.

We therefore investigate the accuracy of replacing W_k with a Dirac-δ function, as was done in Hoekstra et al. (2002) in a different context. If the resulting error were reasonably small, this would allow us to make predictions of b_{2D} and r_{2D} on angular scales were W_k would already be affected by the shot-noise related problems. The quality of this approximation obviously depends on how quickly the bias parameters change with scale. We have therefore obtained a fit to $b(k)$ and $r(k)$ for the red sample that describes well the general features seen in Fig. 8.1. We have then used this fit to compute $\overline{b^2(k)}$ and $\overline{b(k)\,r(k)}$ according to Eqs. (8.21) and (8.22), in one case replacing W_k with $\delta_D\left(k - 4.11/f_K(w)\theta\right)$ (Bartelmann & Schneider 2001), and in the other case using the exact W_k. The results are given in Fig. 8.4. Above $\theta \approx 10'$, where the bias parameters are approximately constant, the accuracy of the approximation is generally better than one percent. For smaller aperture radii, where $b(k)$ and $r(k)$ vary more quickly, the quality of the approximations begins to degrade.

We conclude that predictions of the two-dimensional bias parameters b_{2D} and r_{2D} from the three-dimensional power spectra measured from N-body simulations are difficult both on the smallest and largest scales. For small θ, the filter function W_k is rather broad and therefore is non-negligibly small on scales that are dominated by shot-noise even for aperture radii of the order of arcminutes (the details depend on the galaxy sample under consideration). Unfortunately, replacing W_k with a delta-function to avoid this problem is not a solution, since in particular on small scales the three-dimensional bias and correlation factors vary quickly. A similar problem occurs on very large scales, where $b(k)$ and $r(k)$ can not be computed due to the finite size of
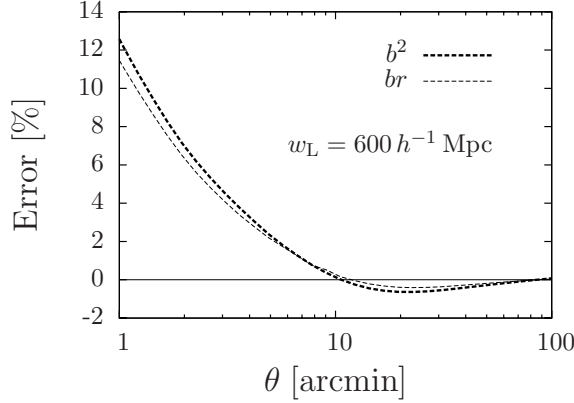
Figure 8.4: *Fractional error of $\overline{b^2}$ and $\overline{br}$ if $W_k(k)$ is approximated by a Dirac delta-function. The input bias factors are fits to the measured $b(k)$ and $r(k)$ for the red sample. All lenses are at $w_L = 600\,h^{-1}$ Mpc.*

the simulation box. In this case, however, the bias parameters are almost constant, so that the approximation $W_k \propto \delta_D\left(k - 4.11/f_K(w)\theta\right)$ can be used without introducing a significant error.

8.5 Estimating the two-dimensional bias parameters

We now turn to the practical aspects of estimating the projected bias parameters using weak gravitational lensing. There are two possible ways to obtain b_{2D} and r_{2D} from our simulation data. The first is to compute estimates of the projected bias parameters for each of the N_r individual realizations and to subsequently average over all realizations (compare to Eqs. 8.35 and 8.36):

$$\hat{b}_1^{(N_r)}(\theta) = \frac{f_b^{1/2}(\theta)}{N_r} \sum_{i=1}^{N_r} \sqrt{\frac{[\mathcal{N}^2(\theta)]_i}{[M_{ap}^2(\theta)]_i}} \,, \tag{8.40}$$

$$\hat{r}_1^{(N_r)}(\theta) = \frac{f_r(\theta)}{N_r} \sum_{i=1}^{N_r} \frac{[\mathcal{N}(\theta)M_{ap}(\theta)]_i}{\sqrt{[\mathcal{N}^2(\theta)]_i\,[M_{ap}^2(\theta)]_i}} \,, \tag{8.41}$$

where $\left[\mathcal{N}^2\right]_i$, $\left[M_{ap}^2\right]_i$ and $\left[\mathcal{N}\,M_{ap}\right]_i$ are the estimates of the aperture statistics for the i-th realization. The second method consists of averaging the aperture statistics over all realizations and computing b and r from these averages:

$$\hat{b}_2^{(N_r)}(\theta) = f_b^{1/2}(\theta) \sqrt{\frac{\{\mathcal{N}^2(\theta)\}_{N_r}}{\{M_{ap}^2(\theta)\}_{N_r}}} \,, \tag{8.42}$$

$$\hat{r}_2^{(N_r)}(\theta) = f_r(\theta) \frac{\{\mathcal{N}(\theta)M_{ap}(\theta)\}_{N_r}}{\sqrt{\{\mathcal{N}^2(\theta)\}_{N_r}\,\{M_{ap}^2(\theta)\}_{N_r}}} \,, \tag{8.43}$$

where we have defined

$$\{A\}_{N_r} = \frac{1}{N_r} \sum_{i=1}^{N_r} [A]_i \,. \tag{8.44}$$

Eqs. (8.40) and (8.41) have the advantage that the errors on $\hat{b}_1^{(N_r)}$ and $\hat{r}_1^{(N_r)}$ can be estimated from the field-to-field variations. On the other hand, $\hat{b}_1^{(N_r)}$ and $\hat{r}_1^{(N_r)}$ can be severely biased
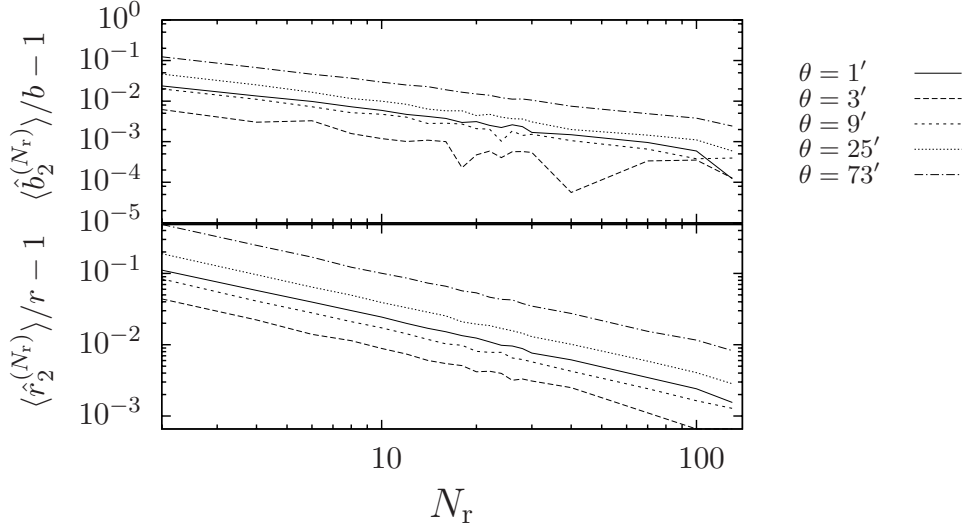
Figure 8.5: *Estimated fractional bias of the estimators $\hat{b}_2^{(N_\mathrm{r})}$ (upper panel, Eq. 8.42) and $\hat{r}_2^{(N_\mathrm{r})}$ (lower panel, Eq. 8.43) as function of the number of realizations N_r. Different curves are for different aperture radii θ.*

due to the non-linear dependence on $\left[\mathcal{N}^2\right]_i$, $\left[M_\mathrm{ap}^2\right]_i$ and $\left[\mathcal{N}\,M_\mathrm{ap}\right]_i$; in fact, they are not even asymptotically unbiased (i.e. $\lim_{N_\mathrm{r}\to\infty}\hat{b}_1^{(N_\mathrm{r})}\neq b_\mathrm{2D}$ and $\lim_{N_\mathrm{r}\to\infty}\hat{r}_1^{(N_\mathrm{r})}\neq r_\mathrm{2D}$). This can be seen by expanding Eqs. (8.40) and (8.41) into Taylor series around the expectation values $\langle M_\mathrm{ap}^2\rangle$, $\langle\mathcal{N}M_\mathrm{ap}\rangle$ and $\langle\mathcal{N}^2\rangle$ and taking the expectation value $\langle\ldots\rangle$ of the resulting expression. To second order, we find

$$\left\langle\hat{b}_1^{(N_\mathrm{r})}\right\rangle = b_\mathrm{2D} - \frac{b_\mathrm{2D}}{8\left\langle\mathcal{N}^2\right\rangle^2}\sigma_{[\mathcal{N}^2]}^2 + \frac{3b_\mathrm{2D}}{8\left\langle M_\mathrm{ap}^2\right\rangle^2}\sigma_{[M_\mathrm{ap}^2]}^2 - \frac{\mathrm{Cov}\left([\mathcal{N}^2]\,[M_\mathrm{ap}^2]\right)}{4b_\mathrm{2D}\left\langle\mathcal{N}^2\right\rangle\left\langle M_\mathrm{ap}^2\right\rangle}f_b + \ldots \quad (8.45)$$

$$\left\langle\hat{r}_2^{(N_\mathrm{r})}\right\rangle = r_\mathrm{2D} + \frac{3r_\mathrm{2D}}{8\langle\mathcal{N}^2\rangle^2}\sigma_{[\mathcal{N}^2]}^2 + \frac{3r_\mathrm{2D}}{8\langle M_\mathrm{ap}^2\rangle^2}\sigma_{[M_\mathrm{ap}^2]}^2 + \frac{r_\mathrm{2D}}{4\langle\mathcal{N}^2\rangle\left\langle M_\mathrm{ap}^2\right\rangle}\mathrm{Cov}\left([\mathcal{N}^2]\,[M_\mathrm{ap}^2]\right)$$

$$- \frac{\mathrm{Cov}\left([\mathcal{N}M_\mathrm{ap}]\,[M_\mathrm{ap}^2]\right)}{2\langle\mathcal{N}^2\rangle^{1/2}\left\langle M_\mathrm{ap}^2\right\rangle^{3/2}}f_r + \frac{\mathrm{Cov}\left([\mathcal{N}M_\mathrm{ap}]\,[\mathcal{N}^2]\right)}{2\langle\mathcal{N}^2\rangle^{3/2}\left\langle M_\mathrm{ap}^2\right\rangle^{1/2}}f_r + \ldots . \quad (8.46)$$

Here, $b_\mathrm{2D} = \sqrt{f_b\left\langle\mathcal{N}^2\right\rangle/\left\langle M_\mathrm{ap}^2\right\rangle}$ and $r_\mathrm{2D} = f_r\left\langle\mathcal{N}M_\mathrm{ap}\right\rangle/\sqrt{\left\langle\mathcal{N}^2\right\rangle\left\langle M_\mathrm{ap}^2\right\rangle}$ are the true values of the bias and correlation factors. The quantities $\sigma_{[\cdot]}^2$ are the variances of $\left[\mathcal{N}^2\right]_i$, $\left[\mathcal{N}\,M_\mathrm{ap}\right]_i$ and $\left[M_\mathrm{ap}^2\right]_i$, respectively. Since the $\sigma_{[\cdot]}^2$ are independent of N_r, the terms depending on the moments of the aperture statistics in $\hat{b}_1^{(N_\mathrm{r})}$ and $\hat{r}_1^{(N_\mathrm{r})}$ do not decrease as the number of realizations increases.

The situation is different for the estimators $b_2^{(N_\mathrm{r})}$ and $r_2^{(N_\mathrm{r})}$. In this case, we find expansions

that are almost identical to Eqs. (8.45) and (8.46),

$$
\left\langle \hat{b}_2^{(N_{\mathrm{r}})} \right\rangle = b_{2\mathrm{D}} - \frac{b_{2\mathrm{D}}}{8\left\langle \mathcal{N}^2 \right\rangle^2} \sigma_{\{\mathcal{N}^2\}}^2 + \frac{3b_{2\mathrm{D}}}{8\left\langle M_{\mathrm{ap}}^2 \right\rangle^2} \sigma_{\{M_{\mathrm{ap}}^2\}}^2
$$
$$
- \frac{\mathrm{Cov}\left(\{\mathcal{N}^2\}\{M_{\mathrm{ap}}^2\}\right)}{4b_{2\mathrm{D}}\left\langle \mathcal{N}^2 \right\rangle\left\langle M_{\mathrm{ap}}^2 \right\rangle} f_b + \dots \ , \tag{8.47}
$$

$$
\left\langle \hat{r}_2^{(N_{\mathrm{r}})} \right\rangle = r_{2\mathrm{D}} + \frac{3r_{2\mathrm{D}}}{8\langle \mathcal{N}^2 \rangle^2} \sigma_{\{\mathcal{N}^2\}}^2 + \frac{3r_{2\mathrm{D}}}{8\langle M_{\mathrm{ap}}^2 \rangle^2} \sigma_{\{M_{\mathrm{ap}}^2\}}^2 + \frac{r_{2\mathrm{D}}}{4\langle \mathcal{N}^2 \rangle\langle M_{\mathrm{ap}}^2 \rangle} \mathrm{Cov}\left(\{\mathcal{N}^2\}\{M_{\mathrm{ap}}^2\}\right)
$$
$$
- \frac{\mathrm{Cov}\left(\{\mathcal{N}M_{\mathrm{ap}}\}\{M_{\mathrm{ap}}^2\}\right)}{2\langle \mathcal{N}^2 \rangle^{1/2}\left\langle M_{\mathrm{ap}}^2 \right\rangle^{3/2}} f_r + \frac{\mathrm{Cov}\left(\{\mathcal{N}M_{\mathrm{ap}}\}\{\mathcal{N}^2\}\right)}{2\left\langle \mathcal{N}^2 \right\rangle^{3/2}\left\langle M_{\mathrm{ap}}^2 \right\rangle^{1/2}} f_r + \dots \ , \tag{8.48}
$$

with the important difference, however, that the $\sigma_{\{\cdot\}}^2$ are now the variances of the *means* of $\left[\mathcal{N}^2\right]_i$, $\left[\mathcal{N}M_{\mathrm{ap}}\right]_i$ and $\left[M_{\mathrm{ap}}^2\right]_i$ over N_{r} realizations, which *do* decrease with the number of realizations ($\propto 1/N_{\mathrm{r}}$). Therefore, the estimators $b_2^{(N_{\mathrm{r}})}$ and $r_2^{(N_{\mathrm{r}})}$ are asymptotically unbiased.

This is illustrated in Fig. 8.5, where we show a rough estimate of the bias of $\hat{b}_2^{(N_{\mathrm{r}})}$ and $\hat{r}_2^{(N_{\mathrm{r}})}$ as function of N_{r} for various aperture radii θ. To obtain these estimates, we fitted the distributions of $\left[\mathcal{N}^2\right]_i$, $\left[M_{\mathrm{ap}}^2\right]_i$ and $\left[\mathcal{N}M_{\mathrm{ap}}\right]_i$ found from the ray-racing simulations of the blue sample with a log-normal distribution function. In doing so, we omit the few cases where one of the aperture measures is negative due to noise. This biases the means of the distributions slightly upwards, which, however, is unimportant for our purpose. By sampling from the fitted distributions, we created a large number of mock samples of the aperture measures for various values of N_{r} (neglecting possible correlations between $\left[\mathcal{N}^2\right]_i$, $\left[M_{\mathrm{ap}}^2\right]_i$ and $\left[\mathcal{N}M_{\mathrm{ap}}\right]_i$ for fixed aperture radius). From these, we have computed the estimators $\hat{b}_2^{(N_{\mathrm{r}})}$ and $\hat{r}_2^{(N_{\mathrm{r}})}$. Finally, the expectation value of these estimators for a given N_{r} was estimated by averaging over their values computed from the different mock samples. The "true" values of $b_{2\mathrm{D}}$ and $r_{2\mathrm{D}}$ are known because we know the means of the fitted log-normal distribution functions. Fig. 8.5 shows that for our setup the bias of the two estimators $\hat{b}_2^{(N_{\mathrm{r}})}$ and $\hat{r}_2^{(N_{\mathrm{r}})}$ is of the order of $\approx 1\%$ for all aperture radii if $N_{\mathrm{r}} \gtrsim 50$. Since we have neglected possible correlations between the different aperture statistics in our numerical experiment, we can therefore not be sure that this bias is negligibly small (i.e. $\ll 1\%$). We therefore experimentally subtract the second order terms in Eqs. (8.47) and (8.48) from our estimates of b and r when using data from the ray-tracing simulations. The effect of this is negligible for both red and blue samples, indicating that the bias is indeed small.

We estimate the variance of the estimators $\hat{b}_2$ and $\hat{r}_2$ for a particular angular bin from the variances of the aperture statistics using standard error propagation, also taking the correlations between $\{\mathcal{N}^2\}$, $\{M_{\mathrm{ap}}^2\}$ and $\{\mathcal{N}M_{\mathrm{ap}}\}$ into account. For that, we expand $\hat{b}_2$ and $\hat{r}_2$ into Taylor series to linear order in $\{\mathcal{N}^2\}$, $\{M_{\mathrm{ap}}^2\}$ and $\{\mathcal{N}M_{\mathrm{ap}}\}$ around their expectation values, square the resulting expression and take the expectation value to obtain the variance. The result for $\hat{b}_2$ is

$$
\mathrm{Var}(\hat{b}_2)(\theta) = \boldsymbol{d}_b^{\mathrm{t}}(\theta)\, \mathsf{C}(\theta)\, \boldsymbol{d}_b(\theta)\ , \quad \text{where}\quad \boldsymbol{d}_b(\theta) = \left(\frac{\partial \hat{b}_2(\theta)}{\partial\left\{\mathcal{N}^2(\theta)\right\}_{N_{\mathrm{r}}}}, \frac{\partial \hat{b}_2(\theta)}{\partial\left\{M_{\mathrm{ap}}^2(\theta)\right\}_{N_{\mathrm{r}}}} \right)^{\mathrm{t}}. \tag{8.49}
$$

Here, the covariance matrix C is given by

$$
\mathsf{C}(\theta) = \begin{pmatrix} \sigma_{\{\mathcal{N}^2\}}^2(\theta) & \mathrm{Cov}\left(\{\mathcal{N}^2\}\{M_{\mathrm{ap}}^2\}\right)(\theta) \\ \mathrm{Cov}\left(\{\mathcal{N}^2\}\{M_{\mathrm{ap}}^2\}\right)(\theta) & \sigma_{\{M_{\mathrm{ap}}^2\}}^2(\theta) \end{pmatrix}. \tag{8.50}
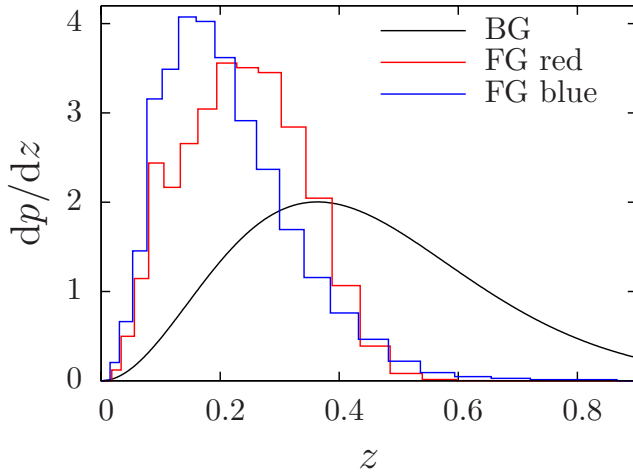$$

Figure 8.6: *Redshift distributions of the blue and red lens samples and the assumed distribution of source galaxies.*

We estimate the variances and covariances of the means of the aperture statistics in Eq. (8.50) by computing $\sigma^2_{[M^2_{\mathrm{ap}}]}$, $\sigma^2_{[\mathcal{N}^2]}$ and $\mathrm{Cov}\left([\mathcal{N}^2][M^2_{\mathrm{ap}}]\right)$ from the field-to-field variations and subsequently dividing by N_{r}. An equation similar to Eq. (8.49) applies to the variance of $\hat{r}_2$.

Finally, we discuss the implications of our findings for realistic survey geometries. In the case that the survey consists of a large number of unrelated telescope pointings (such as in Simon et al. 2007), the discussion of this section applies and the estimators given in Eqs. (8.42) and (8.43) should be used. The situation is different if the survey consists of a single large, contiguous field, as will be the case for the upcoming surveys such as KIDS and Pan-STARRS (Kaiser & Pan-STARRS Collaboration 2005). Breaking up the survey area into several smaller sub-units would result in a degradation of the estimates of the correlation functions and the galaxy-galaxy lensing signal due to the overall smaller number of galaxy pairs in a given angular separation bin. For these surveys, however, the variances and higher moments in the expansions for the estimators $\hat{b}_1$ and $\hat{r}_1$, as given by Eqs. (8.45) and (8.46), are most likely small enough so that the resulting bias becomes negligible.

8.6 Accuracy of f_b and f_r

In order to estimate the projected bias parameters b_{2D} and r_{2D}, the functions $f_b(\theta)$ and $f_r(\theta)$ have to be computed from theory. In Sec. 8.3, we have derived expressions for these functions using the Born approximation. However, as we have seen in Secs. 3.4 and 3.5, higher-order lensing effects can not always be neglected, in particular for volume-limited lens galaxy samples. The resulting error in f_b and f_r leads directly to a bias in the estimates of the bias parameters.

We begin with a validation of the results for the higher-order corrections to the GGL signal given in Sec. 3.4. For this, we have computed $\langle\gamma_{\mathrm{t}}(\theta)\rangle$ for unbiased lens galaxies (i.e. $b = r = 1$, drawn randomly from the dark matter particles), both using full ray-tracing and switching light deflections and non-linear effects off in the ray-tracing code. In Fig. 8.7, we compare the measured ratios $\langle\gamma_{\mathrm{t}}^{\mathrm{Born}}\rangle/\langle\gamma_{\mathrm{t}}^{\mathrm{fullRT}}\rangle$ to the theoretical predictions based on Eqs. (3.128)-(3.130) for three different lens redshifts z_{L}. Within the error bars of the simulations, we find that our theoretical predictions reproduce the behaviour seen in the simulations very well.

We show a similar comparison for the more realistic blue and red lens galaxy samples in Fig. 8.8. Whereas the difference between Born approximation and full ray-tracing is $\lesssim 2\%$ for
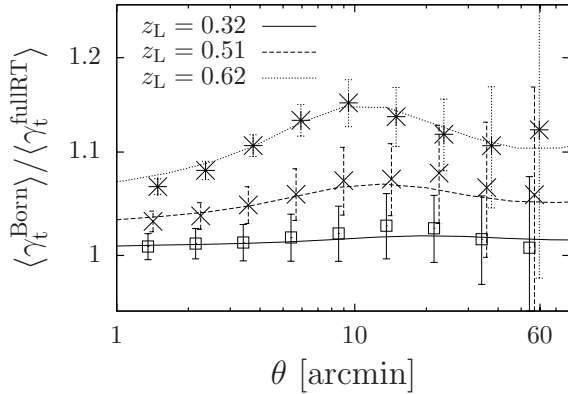
Figure 8.7: *Ratio of the galaxy-galaxy lensing signal $\langle\gamma_t(\theta)\rangle$ computed in Born approximation to the case when higher-order corrections are included for different lens redshifts z_L. The lens galaxies are unbiased ($b = r = 1$) and the sample is volume-limited. Symbols with error bars are simulation results, lines are based on the prediction from Sec. 3.4.*
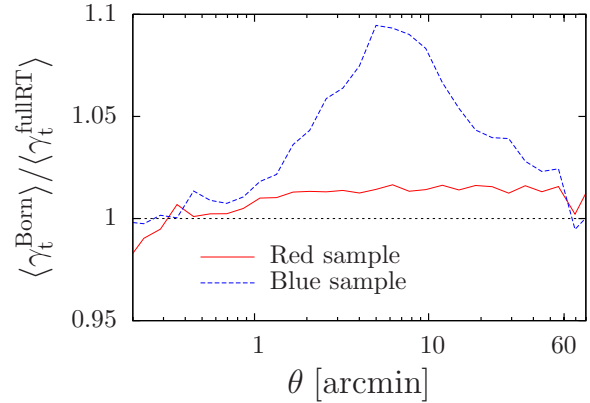
Figure 8.8: *Ratio of the galaxy-galaxy lensing signals $\langle\gamma_t(\theta)\rangle$ computed in Born approximation and using full ray-tracing for the blue and red lens galaxy samples.*
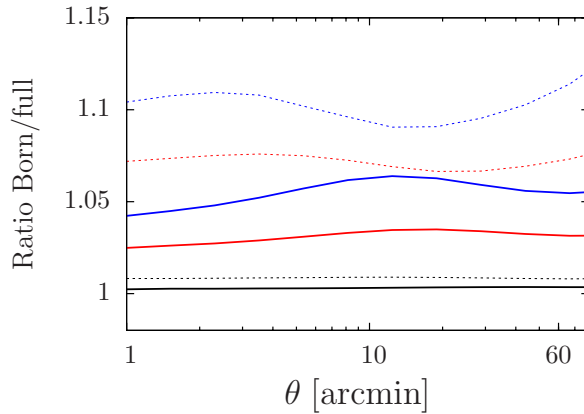
Figure 8.9: *Ratios of the Born approximation predictions to the predictions including the higher-order corrections derived for $\langle\gamma_t(\theta)\rangle$ (solid lines) and $\omega(\theta)$ (dashed lines), respectively. We have assumed a volume-limited lens galaxy sample. The source galaxies are at $z = 0.9$. For the lens redshift distributions we use Eq. (5.17) with $\alpha = 2$; the remaining parameters are given in the legend.*

the red sample, it can be as large as 10% for the blue sample. While these effects can in principle be modelled if the luminosity function of the lens galaxies is known (see Sections 3.4 and 3.5), for now we constrain ourselves to the discussion of volume-limited samples.

As discussed in 3.5, also the angular correlation function, and therefore $\langle\mathcal{N}^2\rangle$, is affected by higher-order effects. In 8.9, we show the ratio of the predictions using the Born approximation to the predictions made including the higher-order correction terms both for the galaxy-galaxy lensing signal and for the angular correlation function. For a given lens redshift distribution, the error made by using the Born approximation for $\omega(\theta)$ is roughly the square of the error for $\langle\gamma_t(\theta)\rangle$, relatively independent of the redshift distribution of the background galaxies. Since the function f_r is proportional to $\sqrt{\langle\mathcal{N}^2\rangle}/\langle\mathcal{N}M_{ap}\rangle$, this means that these errors approximately cancel out. This is not the case for $f_b \propto \langle\mathcal{N}^2\rangle^{-1}$. Since the Born approximation overpredicts the angular correlation function, f_b will be too small, biasing the estimate of b_{2D} low.

In addition to higher-order lensing effects, it has to be taken into account that the functions f_b and f_r depend on the cosmological parameters, even though b_{2D} and r_{2D} (Eqs. 8.35 and 8.36)
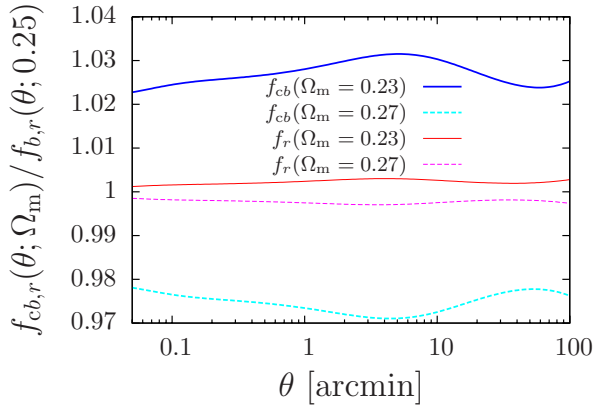
Figure 8.10: *Dependence of f_b and f_r on the matter density parameter Ω_m. Shown are the ratios of $f_{cb}(\theta; \Omega_m) = \left(\Omega_m^{\mathrm{fid}}/\Omega_m\right)^2 f_b(\theta; \Omega_m)$ and $f_r(\theta; \Omega_m)$ to their values for the fiducial cosmology ($\Omega_m^{\mathrm{fid}} = 0.25$).*
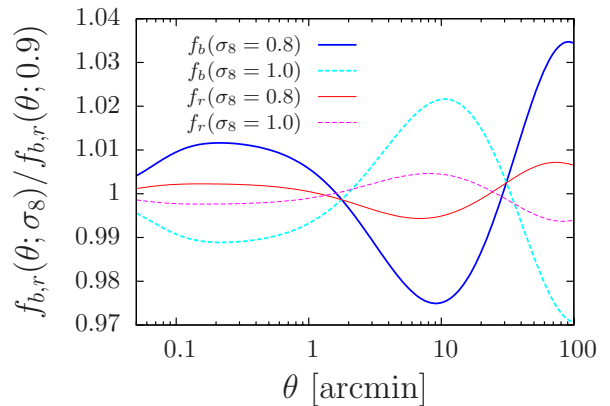
Figure 8.11: *Dependence of f_b and f_r on the power spectrum normalization σ_8. Shown are the ratios of $f_b(\theta; \sigma_8)$ and $f_r(\theta; \sigma_8)$ to their values for the fiducial cosmology ($\sigma_8^{\mathrm{fid}} = 0.9$).*

have been constructed to minimize this dependence (they depend only on ratios of the power spectra). Using the fit formula by Smith et al. (2003) for the power spectrum, we explore the dependence on Ω_m and σ_8 in Figs. 8.10 and 8.11, respectively, with the additional constraint that the Universe remain flat. We plot the fractional changes with respect to our fiducial cosmology, $\Omega_m^{\mathrm{fid}} = 0.25$ and $\sigma_8^{\mathrm{fid}} = 0.9$. Since f_b is by construction proportional to Ω_m^2 through its dependence on $h_{\kappa\kappa}$ (Eq. 8.37)[1], we use the re-scaled function $f_b' = \left(\Omega_m^{\mathrm{fid}}/\Omega_m\right)^2 f_b$ for the comparison. The figures show that the two functions are indeed remarkably stable with respect to changes of the cosmological parameters; for the range of values we consider for Ω_m and σ_8, the difference never exceeds 4%. The oscillations seen in Fig. 8.11 come about since by varying σ_8 one does not only alter the amplitude of the power spectrum but also its shape by changing the wave number at which the transition between the linear and non-linear regimes occurs. These findings imply that uncertainties of the cosmological parameters will have little practical impact on the measured projected bias and correlation parameters.

8.7 Comparison to the ray-tracing simulations

In the previous sections, we have discussed in isolation several systematic effects that can have an influence on the accuracy with which the projected bias parameters b_{2D} and r_{2D} can be measured. To investigate their combined effect, we now compare the projected bias parameters measured from our ray-tracing simulations to the predictions based on the three-dimensional bias parameters.

In Fig. 8.12, we show the results for the red and blue samples, as well as for both samples joined (i.e. using all lens galaxies with $r < 20$). For this plot, we did not add shape noise to the background galaxies, for which we assume the redshift distribution given by Eq. (5.17) with $\alpha = 2.0$, $\beta = 1.5$, $z_0 = 0.3$ and a sharp cut-off at $z = 0.9$ (the maximum redshift of the ray-tracing simulations), and a number density of $n = 15/\mathrm{arcmin}^2$. As before, we use 128 realizations of a $2° \times 2°$ field.

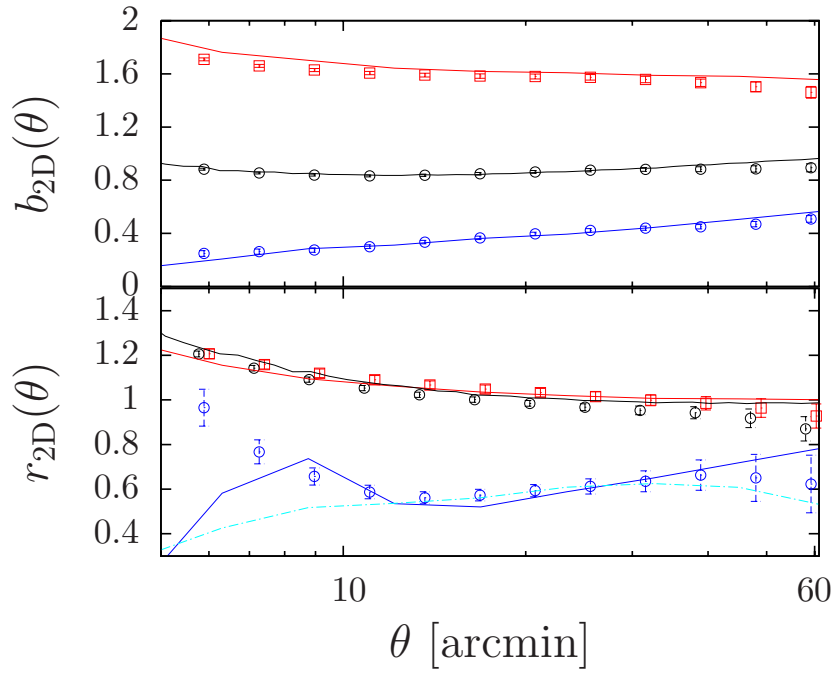[1]Note that this dependence cancels out again when computing b_{2D}

Figure 8.12: *Projected bias and correlation factors for the red, blue and joint lens samples (red, blue and black solid lines and symbols, respectively). The predictions based on the three-dimensional bias parameters (approximating W_k by a delta-function) are shown as solid lines; for the correlation factor of the blue sample, in addition the prediction using the exact W_k is shown as light blue, dashed line. Symbols are the results from the weak lensing technique. No shape noise was added to the background galaxies.*
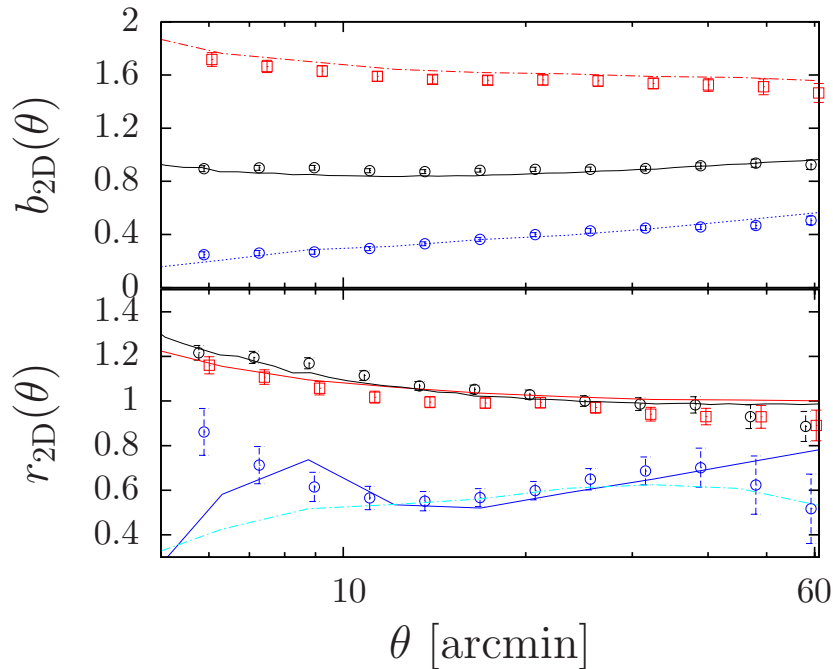
Figure 8.13: *Same as Fig. 8.12, but including shape noise ($\sigma_\epsilon = 0.3$).*

The bias factor $b_{2\mathrm{D}}$ measured from the ray-racing simulations using the estimator $\hat{b}_2$ (Eq. 8.47) agrees with the prediction from the three-dimensional bias parameters within $\approx 3\%$, even though $b_{2\mathrm{D}}$ is slightly biased low. A similar accuracy is achieved for the correlation factor $r_{2\mathrm{D}}$; also here, slight systematic deviations from the predictions are visible.

Note that we only plot the predictions for $b_{2\mathrm{D}}$ and $r_{2\mathrm{D}}$ obtained by replacing the filter W_k in Eqs. (8.21) and (8.22) with a delta-function to avoid the problems caused by the finite width of W_k (as discussed in Sec. 8.4). Comparing these approximate predictions to the ones obtained using the exact form of W_k (Eq. 8.23) in the limited range of scales where shot noise in P_{g} and the finite size of the simulation volume are unimportant, we find that in general the delta-function approximation of W_k is very good. The only exception is the correlation factor of the blue sample, which varies rapidly at $\theta \approx 10'$. In this case we also show the exact prediction (i.e. using the exact form of W_k) as dot-dashed light blue line in addition to the approximate one (dark blue solid line). In the range where this can be computed accurately ($10' \lesssim \theta \lesssim 30'$), the agreement to the estimates from the weak lensing method is very good. On smaller and larger angular scales, W_k picks up significant contributions from wave-numbers where the galaxy power spectrum and thus the three-dimensional bias parameters cannot be estimated reliably.

As can be seen from Fig. 8.13, where we now include shape noise for the background galaxies ($\sigma_\epsilon = 0.3$ for each ellipticity component), intrinsic ellipticities constitute only a small part of the total error budget for a survey of the size considered here.

Besides the effects discussed in the previous section, the systematic deviations seen in Figs. 8.12 and 8.13 can be explained by two further factors. The first is that we might slightly underestimate the size of the error bars by using Eq. (8.49), which relies on the linearizability of the estimators $\hat{b}_2$ and $\hat{r}_2$. This suspicion is supported by the observation that the field-to-field variances of the estimators $\hat{b}_1$ and $\hat{r}_1$ (Eqs. 8.40 and 8.41) yield error estimates that are indeed larger by a factor of ≈ 1.5. However, it is not clear how to relate the variances of these biased estimators to the variances of $\hat{b}_2$ and $\hat{r}_2$.

The second possible (and probably most important) reason for the deviations is the fact that the weak lensing estimation of $b_{2\mathrm{D}}$ and $r_{2\mathrm{D}}$ requires the computation of the aperture statistics from the measured GGL signal $\langle \gamma_{\mathrm{t}} \rangle$, the angular correlation function ω and the shear correlation functions $\xi_\pm$. However, the fact that the latter can only be estimated above a certain minimal angular scale introduces a bias in the estimates of the aperture statistics, since the integrals in Eqs. 8.19, 8.17 and 3.93 extend down to zero angular separation. In our simulation, this is not possible because we do not find pairs of galaxies with arbitrarily small angular separation due to the finite number of lens and background galaxies. In practice, further issues arise such as the finite size of the observed galaxies.

The implications of this have been investigated in detail for $\langle M_{\mathrm{ap}}^2 \rangle$ by Kilbinger et al. (2006). We extend this study to the cases of $\langle \mathcal{N}^2 \rangle$ and $\langle \mathcal{N} M_{\mathrm{ap}} \rangle$ by computing the aperture statistics from the theoretical predictions of $\omega(\theta)$ and $\langle \gamma_{\mathrm{t}}(\theta) \rangle$ for unbiased galaxies; since these are known also for $\theta = 0$, we can compare the case where the lower integration limits are zero to the case with a finite small-scale cutoff. We find a similar susceptibility of $\langle \mathcal{N}^2 \rangle$ and $\langle \mathcal{N} M_{\mathrm{ap}} \rangle$ to a small scale cut-off as for the aperture mass dispersion. This is illustrated for the case of $\langle \mathcal{N}^2 \rangle$ in Fig. 8.14, where we plot the error for various values of the small scale cutoff θ_{min} that roughly correspond to the limitations found using our mock galaxy catalogs. The aperture dispersion is biased low on all relevant angular scales, eventually even becoming negative. A similar effect is observed for $\langle \mathcal{N} M_{\mathrm{ap}} \rangle$. However, θ_{min} is much smaller in this case because the GGL signal can be estimated reliably down to much smaller angular scales. The reason for this is that we have many more background galaxies than lens galaxies, and for the computation of $\langle \gamma_{\mathrm{t}} \rangle$
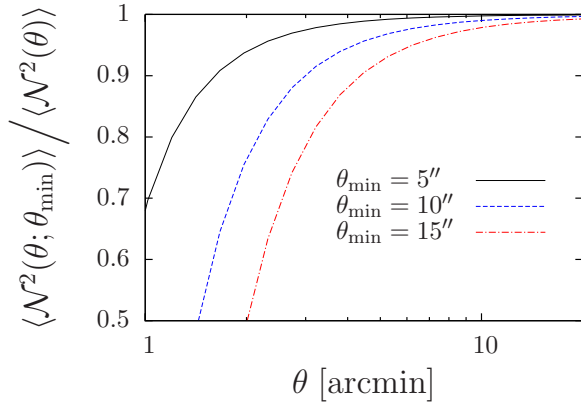
Figure 8.14: *Ratios of $\langle \mathcal{N}^2 \rangle$, computed from the angular correlation ω function using Eq. (8.17), where ω was set to zero below $\theta_{\min}$, to the case without cut-off.*
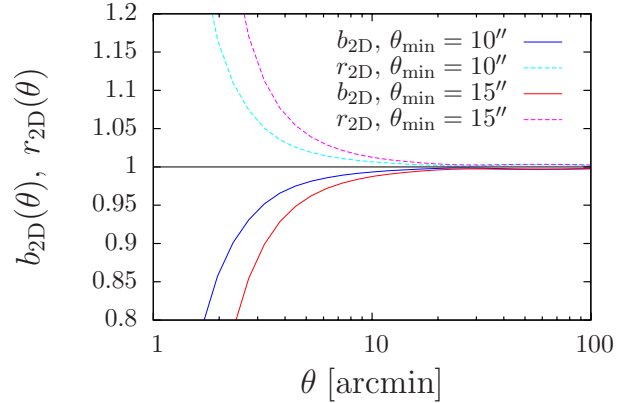
Figure 8.15: *The two-dimensional bias parameters of unbiased galaxies, computed with a small scale cut-off at $\theta_{\min} = 10''$ and $15''$ for the angular correlation function and $\theta_{\min} = 5''$ for $\langle \gamma_t \rangle$.*

we form foreground-background pairs instead of foreground-foreground pairs as for the angular clustering.

We explore the impact of these biases on the projected bias and correlation parameters in Fig. 8.15, choosing $\theta_{\min} = 10''$ or $\theta_{\min} = 15''$ for the angular clustering and $\theta_{\min} = 5''$ for the galaxy-galaxy lensing. Furthermore, we have assumed unbiased lens galaxies with $b = r = 1$. The Figure shows that the bias factor can be biased low by several percent on scales of $\approx 10''$. The opposite is true for the correlation factor, which is found to be consistently too large. These findings indeed explain some of the qualitative features of the systematic deviations in Figs. 8.12 and 8.13, in particular the rapid increase of $\hat{r}_2$ for the blue sample on small angular scales.

The results of splitting the lens samples into low-redshift ($z \leq 0.2$) and high-redshift ($z > 0.2$) subsamples (again not including shape noise) are shown in Figs. 8.16 and 8.17, respectively. The aforementioned biases become even more apparent here due to the lower number density of lens galaxies in each sample; the bias factor of the red sample is now low by as much as $\approx 7\%$, and a similar trend can also be seen on small scales for the joint samples. The estimates of the projected correlation factors b_{2D} degrade particularly strongly. In addition, for the blue sample a reliable prediction from the three-dimensional bias parameters becomes impossible on scales below $\theta \approx 10'$.

We now address the question whether a possible evolution of the bias parameters with redshift can be detected using the weak lensing method. In Fig. 8.18, we plot the predictions for b_{2D} and r_{2D} for the various low- and high-redshift lens samples. While the red lens galaxies are more strongly biased at low redshifts, the opposite is true for the blue lenses. Accordingly, the bias factor of the joint sample does not show any significant time evolution. The correlation factors show little changes on large scales; the increase of r_{2D} with z seen for the red and joint samples is most likely explained by the increasing importance of the shot noise correction for the galaxy power spectrum at higher redshifts (see Sec. 8.4). Since the amplitudes of the evolutionary effects on the bias are similar for red and the blue galaxies, we will in the following focus only on the red sample. We present our results for a survey of with a total area of $128 \deg^2$, composed of 32 fields of the size of $2° \times 2°$, in Fig. 8.19. The survey is thus roughly comparable to currently ongoing projects such as the CFHTLS. The evolution of the bias factor b_{2D} can be detected at high significance on angular scales larger than $\approx 5'$. For smaller θ, systematic effects, most likely

due to the small scale cut-off of the angular correlation function, begin to dominate. Since these are the only scales where the correlation factor displays appreciable time evolution, no definite detection can be made for $r_{2\mathrm{D}}$.

8.8 Summary

In summary, we can say that the ability of the weak lensing method explored in this chapter to measure the projected bias parameters strongly depends on the lens sample selection and the design of the weak lensing survey. We have identified several sources of bias:

- If the number of independent fields or the size of a contiguous survey is small, both estimators for the projected bias parameters (Eqs. 8.40, 8.41 and 8.42, 8.43) are biased; we have estimated that the bias for $\hat{b}_2$ and $\hat{r}_2$ is small once the number of independent fields exceeds ≈ 50.

- In order to compute $b_{2\mathrm{D}}$ and $r_{2\mathrm{D}}$, the cosmology-dependent functions f_b and f_r have to be known. We have shown that these are rather insensitive to changes of the cosmological parameters; the present-day uncertainty about the values of Ω_{m} and σ_8 leads to errors of maximally 5%.

- The functions f_b and f_r are usually computed using the Born approximation. We find that the error caused by neglecting higher-order corrections tends to cancel out for f_r, whereas f_b is generally underestimated. For the red sample, this amounts to a bias of $\lesssim 1\%$ in $b_{2\mathrm{D}}$, whereas for the blue sample a bias of $\approx 4\%$ has to be expected.

- Probably the most important source of systematic effects on small angular scales is the fact that the angular correlation function cannot be measured below a certain angular separation due to the lack of close lens-lens pairs. This effect depends on the number density of lens galaxies and the strength of the angular clustering of the lens galaxies. A solution to this problem could be provided by the so-called ring statistics (Schneider & Kilbinger 2007), which have been originally developed to cleanly separate between cosmic shear E- and B-modes if the shear correlation functions are known only on a finite interval. These statistics can be generalized to the case of angular clustering and galaxy-galaxy lensing and could be used replace the aperture statistics. It remains to be seen, however, whether the ring statistics are sufficiently good proxies of the projected power spectra and how localized the resulting projected bias parameters will be.

- A related problem is encountered when predicting the projected bias parameters from the three-dimensional bias parameters measured in the N-body simulation. On small scales, this is limited by shot noise in the lens galaxy power spectrum, on large scales by the finite size of the simulation volume. For this reason, the smoothing of $b(k)$ and $r(k)$ with the filter W_k cannot be carried out on all scales of interest; instead, W_k has to be approximated with a delta-function.

Given this list of possible difficulties, it is surprising that the weak lensing method yields $b_{2\mathrm{D}}$ and $r_{2\mathrm{D}}$ with an accuracy of a few percent, provided the full lens samples are used. The situation deteriorates quickly of the samples are subdivided further, in particular for the blue sample due to its low clustering amplitude. However, when splitting the red sample in two redshift bins,
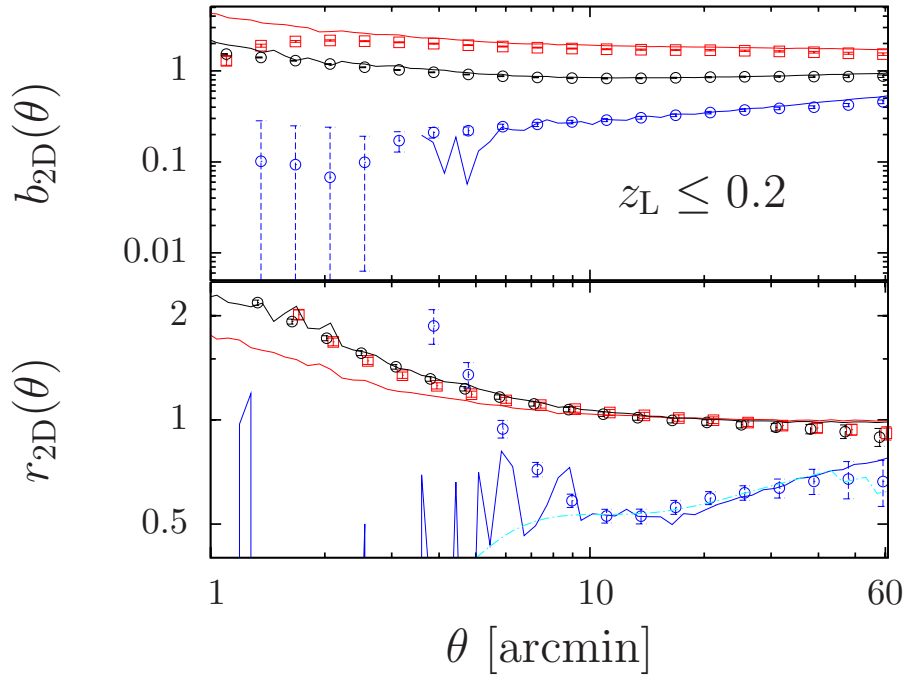
Figure 8.16: *Same as Fig. 8.12, but only for the low-redshift subsamples ($z \leq 0.2$). No shape noise was added to the background galaxies. Note that, unlike before, a logarithmic scale is used for the vertical axes.*
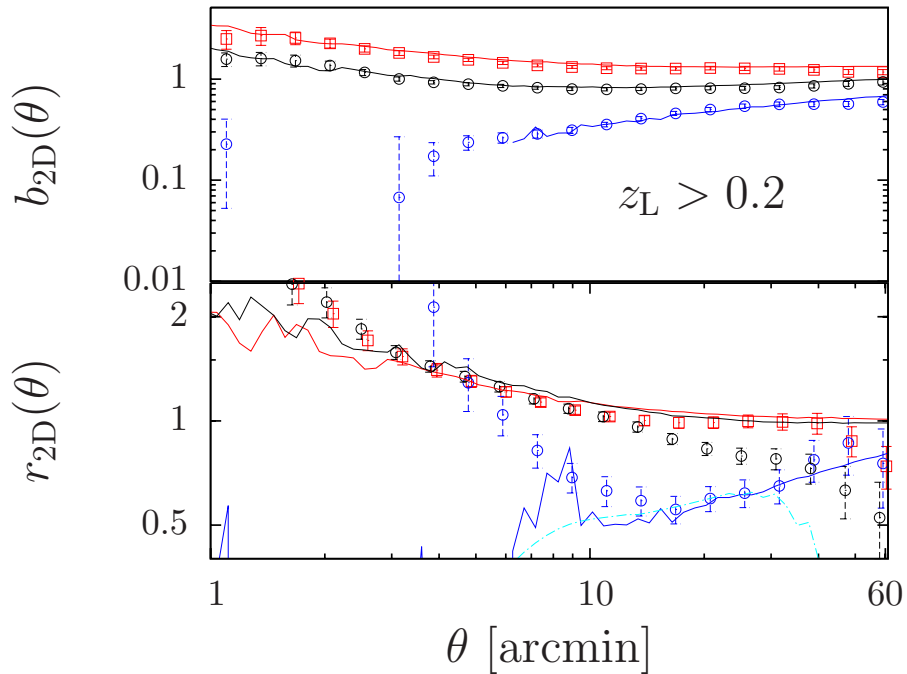

Figure 8.17: *Same as Fig. 8.12, but only for the high-redshift subsamples ($z > 0.2$). No shape noise was added to the background galaxies. Note that a logarithmic scale is used for the vertical axes.*
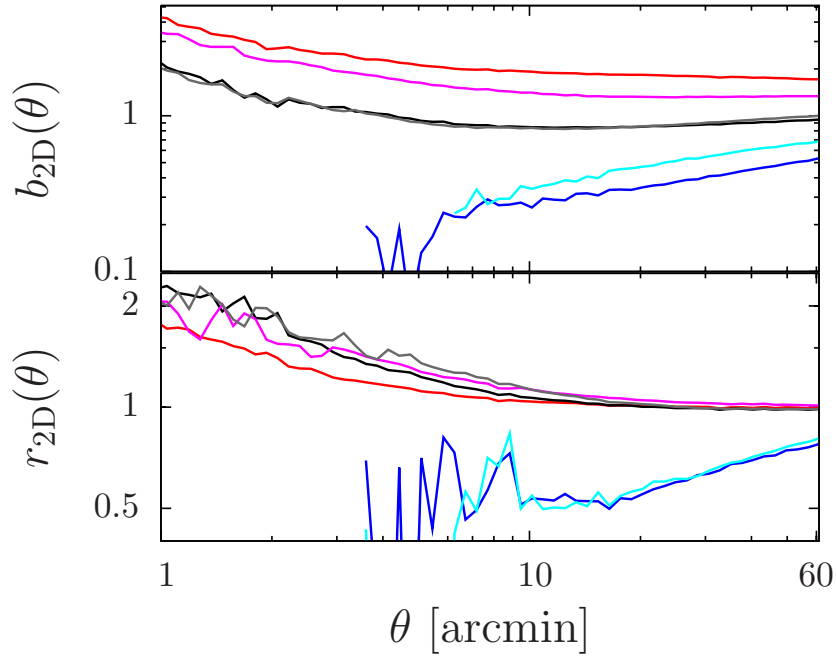
Figure 8.18: *Redshift evolution of the projected bias parameters as predicted from three-dimensional bias parameters. The low redshift subsamples for red, blue and all lens galaxies are given as red, blue and black lines, respectively, whereas the high-redshift samples are given as magenta, light blue and grey lines. Note that a logarithmic scale is used for the vertical axes.*
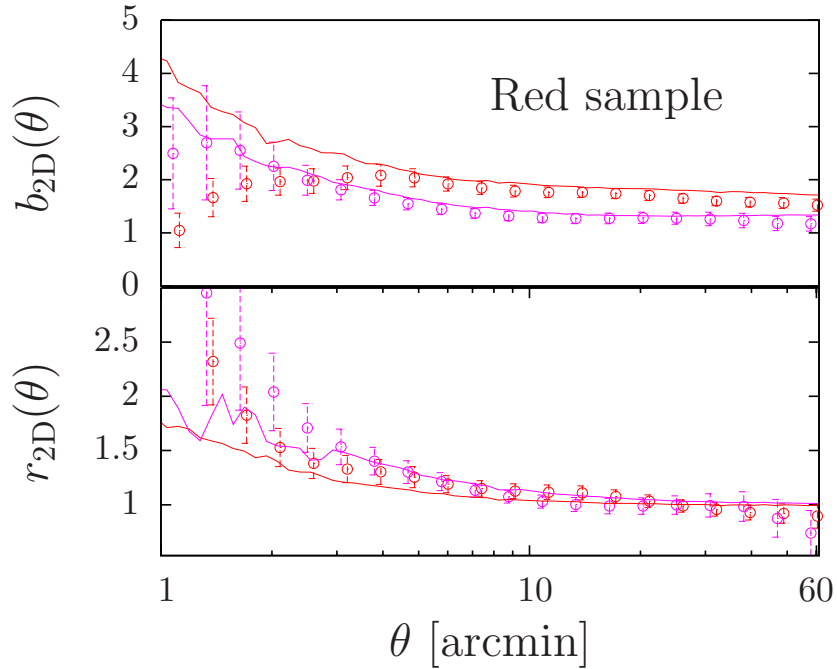

Figure 8.19: *Redshift evolution of the projected bias parameters for the red lens sample. Symbols show the weak lensing estimates, lines the prediction from the three-dimensional bias parameters (low-z subsample: red, high-z subsample: magenta).*

the accuracy is still sufficient to significantly detect the evolution of the bias parameter using a survey that is roughly comparable to the CFHTLS.

In spite of this, we argue that in order to properly account for the issues listed above, observational measurements of b_{2D} and r_{2D} should be compared to the results of ray-tracing simulations that closely mimic the data reduction process, and not to the predictions based on the three-dimensional bias parameters, even though they are much simpler to obtain.

9 Summary and Conclusions

Chapter 2: The Cosmological Standard Model

In this chapter, we have reviewed the the basics of the cosmological standard model and the structure formation in the Universe. According to the standard model, only about four percent of the matter-energy content of the Universe is in the form of normal, baryonic matter. There is strong evidence that more than a fifth of the energy budget is constituted by dark matter, which does not interact via the electro-magnetic interaction and is only noticeable through its gravitational effects. Furthermore, the observed current accelerated expansion of the Universe is best explained by the presence of a component with negative pressure, the so-called dark energy, accounting for the remaining $\approx 75\%$ of the Universe's energy content. The total energy density is very close to the critical density, implying that the Universe has a nearly flat geometry.

The Universe appears homogeneous only when averaged over sufficiently large volumes. On length scales below a few hundred Mpc, a wealth of structures, such as filaments, galaxy clusters and galaxies, exists. These structures have evolved from initially very small perturbations of the almost homogeneous early Universe by gravitational collapse. Dark matter is the driving agent of structure formation; baryons initially follow the dark matter distribution and later condense at the center of the potential wells to form stars and galaxies. We discussed the most important aspects of the theory of structure formation and its implications for the statistics of the density field, which can be described as a homogeneous and isotropic random field. Its statistical properties are usually characterized by the n-point correlation functions or, equivalently, by its polyspectra. We discussed in particular the power spectrum, which is the most important and best understood representative of these statistics.

Chapter 3: Gravitational Lensing

In chapter 3, we reviewed the formalism of gravitational lensing for the cases of an isolated thin lens and of continuous light deflection by the large-scale structure in the Universe. In the latter case, the gravitational tidal field of the matter distribution imprints a weak, systematic distortion pattern (shear field) on observed images of background galaxies. To first order, this makes the images of intrinsically round sources having elliptically shaped isophotes. Even though for individual galaxies this signal is much smaller than the noise caused by the intrinsic galaxy shapes, it can be recovered using statistical methods, assuming that the intrinsic orientations of the galaxies are random. We discussed two ways to extract cosmological information from this.

In the first approach, called cosmic shear, one studies the correlations between the ellipticities of different background galaxies. This allows one to infer the statistical properties of the matter distribution that causes the distortions. We discuss the decomposition of the shear field into E- and B-modes. While it can be shown that gravitational lensing does not create a B-mode in the case of a single, thin lens, this is no longer true for lensing by the large-scale structure. We gave expressions for both the E- and B-mode power spectra of the cosmic shear distortion field and related these quantities to the power spectrum of the matter distribution.

A second way to study the matter distribution in the Universe using weak lensing is galaxy-galaxy lensing. Each foreground galaxy leads to a distortion of the images of galaxies in the

background. This effect is too weak to be detectable for individual galaxies; nonetheless, by stacking the signal of many lens galaxies the noise due to the intrinsic shapes of the background objects can be beaten down. The resulting azimuthally averaged shear profile can be used to study the average mass distribution associated with the lens galaxies. This galaxy-galaxy lensing signal is also a measure of the cross-correlation between galaxies and the underlying dark matter distribution and can be predicted from the cross-power spectrum of galaxies and dark matter. Usually, this is done in the Born approximation, assuming that the deflections of the light rays emitted by the background galaxies caused by the large-scale structure are negligible. We derived correction terms to this approximation that take into account the magnification effect: gravitational lensing by a given lens galaxy magnifies nearby patches of the sky, without changing the number of higher-redshift lens galaxies in that patch. Therefore, close to the lens, fewer lens galaxies at higher redshifts are observed than on average, leading to a decrease of the galaxy-galaxy lensing signal. On the other hand, lensing also boosts the flux of more distant lens galaxies and therefore, in a flux-limited lens galaxy sample, more high-redshift lenses are above the detection threshold. The relative importance of these effects depends on the lens redshifts and their luminosity function; for a volume-limited lens sample, the galaxy-galaxy-lensing signal changes by more than 5% for lenses with a mean redshift $\bar{z} \gtrsim 0.4$. We also discussed the impact of the magnification effect on the observed angular clustering of galaxies. We found that the error made by using the Born approximation is slightly larger than in the case of galaxy-galaxy lensing.

Chapter 4: Simulations of Structure Formation

In Chapter 4, we described numerical methods to solve the Vlasov-Poisson system of equations that governs the formation of dark matter structures in the Universe. The basic idea of these N-body simulations is to sample the six-dimensional phase space with pseudo-particles each representing the mass contained in a particular phase space cell. The distribution of particles is then evolved in time using their equations of motion. We described the basic Particle-Mesh (PM) algorithm and several of its descendants, such as the P^3M and TreePM algorithms, which significantly improve on the small-scale resolution.

A general feature of such simulations is the formation of self-bound dark matter structures, so-called halos. We described several algorithms to identify these structures and discussed their relative merits and shortcomings. The halo catalogs obtained in this way can be used to build semi-analytic models of galaxy formation, assuming that galaxies reside at the centers of halos or subhalos. These models use simple, but physically motivated recipes for the complex baryonic physics involved, such as the cooling and heating of gas, the formation of stars, energy injection by supernovae and active galactic nuclei, to follow the build-up of galaxies in the simulation. We described in some detail the model by De Lucia & Blaizot (2007) that we have used in this thesis for our study of galaxy-galaxy lensing.

Finally, we briefly introduced the N-body simulations used in this thesis: the Millennium Simulation, which is one of the largest and most current dark-matter-only simulations available, and in addition a set of ten smaller simulations which we have run using the publically available TreePM code `GADGET-2` to study the non-Gaussianity of the cosmic shear likelihood (Chapter 6).

Chapter 5: Ray-Tracing Simulations

Simulations of weak gravitational lensing are performed using the multiple-lens plane algorithm: the matter distribution along the line of sight from an observer to a distant source is divided into a number of redshift slices which are thin enough (typically $\approx 100 \, h^{-1} \, \mathrm{Mpc}$) for cosmic evolution to be negligible during the time a light ray needs to traverse the slice. These slices are projected onto lens planes, and it is assumed that light deflections only occur at these planes. Light rays are then shot through the array of lens planes, using recursion relations for their positions at the planes and for the Jacobian matrix of the lens mapping from the observer to each plane.

We discussed two implementations of this algorithm, which differ in the way the matter distribution along the line of sight is constructed from the outputs (snapshots) of an N-body simulation. The first, suitable for small N-body simulation volumes, simply uses the complete N-body simulation volumes at different times for the redshift slices. Repetition of structure along the line of sight is avoided by applying random rotations and translations to the particle distribution in each snapshot. A different approach is necessary for the Millennium simulation due to the large size of the simulation box. In this second implementation, the line of sight is supposed to be tilted with respect to the faces of the simulation volume, and the thickness of the redshift slices is chosen to be only a fraction of the box size. A careful choice of the line of sight avoids seeing the same structures repeatedly and at the same time retains the periodicity of the matter distribution on the lens planes. In this case, however, special care has to be taken to avoid slicing through dark matter halos when constructing the lens planes. We also described an algorithm for obtaining the lensed positions of galaxies from semi-analytic models of galaxy formation, combining for the first time realistic galaxy catalogues with a complete ray-tracing code.

We applied our code to study the accuracy of predictions of the convergence power spectrum using the widely-used fitting formulae for the three-dimensional matter power spectrum by Peacock & Dodds (1996) and Smith et al. (2003) as well as the halo model. We found that the fitting formulae significantly underpredict the small-scale power found in the ray-tracing simulations, which undermines their use for deriving constraints on cosmological parameters. Furthermore, we compared the cosmic shear B-mode signal predicted using the formalism in Chapter 3 to the B-mode measured in the simulations. We found that both agree within a factor of ≈ 2, where the origin of the discrepancy remained unclear. However, the order-of-magnitude agreement confirms that the B-mode will be undetectable in any near-future weak lensing survey.

Chapter 6: Weak Lensing and Statistics

In this Chapter, we addressed several aspects of the statistical analysis of cosmic shear data. Constraints on cosmological parameters are usually derived from weak lensing surveys by measuring functions of the shear power spectrum, in our case the shear correlation functions $\xi_\pm$. A best fitting model prediction is then determined either using a Bayesian approach or the maximum likelihood technique. In both approaches, the likelihood function has to be specified. It is commonly assumed that the likelihood is a multivariate Gaussian distribution. In this case, the inverse of the covariance matrix of the shear correlation functions has to be obtained. We discussed the case where it is estimated either from the data themselves (from N independent observations of the correlation functions) or from simulations (from N ray-tracing realizations). We proved analytically that the covariance matrix is singular if the number of bins p for which

the correlation functions have been measured exceeds N (if the mean is known) or $N-1$ (in the more realistic case where the mean of $\xi_\pm$ is unknown). We noted that even if $p < N$, the inverse of the estimated covariance matrix is a biased estimate of the inverse population covariance. If uncorrected, this bias can lead to a severe underestimation of the size of confidence/credible regions. We reported on the existence of a simple correction for this bias, which is valid for $p < N$ provided the measurement errors are Gaussian and the observations are indeed independent. We conducted Monte-Carlo experiments to verify this correction formula; we found that it is also sufficiently accurate in the case where the covariance matrix has been estimated from a single weak lensing observation by bootstrapping (in this case, the bootstrapping realizations are not independent). It has been suggested in the literature that a pseudo-inverse of the covariance matrix can be obtained in the case of $p > N$ using the Singular Value Decomposition. We found that the pseudo-inverse can be severely biased in a way that cannot be easily corrected for.

We then proceeded to study the validity of the assumption that the likelihood function for the shear correlation functions is Gaussian. Since the likelihood is by definition the probability distribution of measuring a certain realization of $\xi_\pm$, we developed and tested a method to estimate the likelihood function from a large set of ray-tracing simulations. Our approach, therefore, is to reduce the dimensionality by finding a linear transformation $\breve{\xi}_\pm = \mathsf{A}\xi_\pm$ that makes the components of the transformed $\breve{\xi}_\pm$ approximately statistically independent. In this case, the likelihood factorizes into the product of p one-dimensional probability distributions which can be estimated using standard techniques. Even though the method fails to fully remove statistical dependencies, we found by testing against an alternative method (Projection Pursuit Density Estimation) that this does not affect our conclusions. As a case study, we applied the method to 9600 ray-tracing realizations of a field similar to the Chandra Deep Field South. We found that the likelihood function is indeed significantly non-Gaussian; most strikingly, this leads to credible regions that are considerably smaller than in the Gaussian case.

Currently, there is considerable effort to obtain reliable theoretical predictions of the covariance matrix of cosmic shear (e.g. Cooray & Hu 2001b; Semboloni et al. 2007; Takada & Jain 2008; Pielorz 2008), assuming that the likelihood is Gaussian. Taking into account that the shear field is a non-Gaussian random field, these authors find that the credible regions *increase* compared to the case where the shear field is assumed to be a Gaussian random field. We argued that in order to accurately study the ability of future surveys to constrain cosmological parameters, the effect of the non-Gaussianity of the likelihood function has to be taken into account, since it might partially or even fully compensate this increase.

Finally, using the non-Gaussian likelihood, we re-derived constraints on the power spectrum normalization σ_8 from the cosmic shear data obtained from the Chandra Deep Field South by Schrabback et al. (2007), who originally had found an extraordinarily low value. Using the ray-tracing simulations, we also quantified possible systematic effects due to the criteria that were used to select the CDFS. We estimated that the requirement that the CDFS contain no significant X-ray source (assuming a flux limit similar to that used for the ROSAT All-Sky Survey) can bias σ_8 low by $\approx 3-5\%$, whereas demanding the absence of a significant extragalactic source might lead to a decrease of $\approx 10\%$. This analysis leads to an estimate of $\sigma_8 = 0.68^{+0.09(\mathrm{stat.})\ +0.1(\mathrm{sys.})}_{-0.16(\mathrm{stat.})}$.

Chapter 7: Galaxy-Galaxy Lensing with the Millennium Simulation

Galaxy-galaxy lensing allows one to infer the average excess surface mass density around lens galaxies. In Chapter 7, we studied the GGL signal as a function of the mass of the lens galaxy

halo and its environment. We found that on small scales, the signal is dominated by the lens galaxy halo. If a significant fraction of lens galaxies are satellite galaxies of a group or cluster halo, the host halo contributes most to the surface mass density on intermediate scales. Finally, on the largest scales, large-scale structure correlated with the lenses dominates the galaxy-galaxy lensing signal. The covariance matrix of the GGL signal displays a similar three-part structure. A simple halo model based on these three components indeed recovers well the properties of the galaxy and host halos when fitted to the GGL signal of lens galaxies that lie in a narrow mass interval. The model fails, however, for more heterogeneous lens samples, since it does not account for the considerable scatter in the halo properties.

Currently, the best measurements of galaxy-galaxy lensing have been obtained using SDSS data (Sheldon et al. 2004). We obtained the GGL signals from our simulations and the galaxy formation model by De Lucia & Blaizot (2007) for lens galaxy samples selected in the same way as in Sheldon et al. (2004). The comparison to the SDSS data revealed a significant discrepancy in the blue filter bands, where the simulations underpredict the signal for luminous blue galaxies. This is most likely due to the insufficient modelling of dust attenuation in the semi-analytic model. The high value of $\sigma_8 = 0.9$ in the Millennium Simulation is expected to lead to an overprediction of the galaxy-galaxy lensing signal by $\approx 10\%$. That we did not observe this effect is probably due to the fact that the galaxy formation model has been tuned to match the properties of observed galaxies, which makes a quantitative comparison of simulation and observation difficult. We also investigated the influence of baryons on the GGL signal using a semi-analytic modelling of the stellar mass (Hilbert et al. 2008b). In our simulations, this leads to a steepening of the excess surface mass density, for which we found tentative evidence when comparing to the SDSS measurements.

Chapter 8: Bias and Correlation Factors from Weak Lensing

Even though dark matter is the driving agent of cosmological structure formation, the distribution of galaxies is not the same as the distribution of the dark matter; moreover, it depends on the properties of the galaxy sample under consideration. This is described by the concept of galaxy bias: the density field of galaxies is a (not necessarily deterministic) function of the dark matter density. In the linear stochastic bias prescription, it is assumed that the power spectra of dark matter and galaxy distribution and their cross power spectrum are linearly related. Using weak lensing, corresponding aperture statistics, i.e., averages over wave number and redshift of the projected power spectra, can be obtained. Therefore, it provides means to estimate projected versions of the linear stochastic bias parameters.

In Chapter 8, we investigated how accurately the bias parameters can be predicted from N-body simulations and measured from realistic data sets. The prediction from the three-dimensional power spectra measured from the N-body simulation is limited by shot noise in the galaxy power spectrum on small scales and by the finite simulation box size on large scales. This prevents the smoothing of the three-dimensional bias parameters over the wave number. Replacing the corresponding smoothing kernel with a delta-function is a reasonable approximation for realistic lens samples only on scales larger than $\approx 10'$.

We studied two estimators for the projected bias parameters obtained from the weak lensing method, suitable for surveys consisting of a single contiguous field or several independent fields. Neither estimator is unbiased; however, the bias decreases with increasing field size or with increasing number of independent fields, respectively. Several other systematic effects can affect the accuracy of the weak lensing method. To obtain the projected bias parameters, the aperture

statistics for unbiased lens galaxies have to be computed from theory. This causes the estimates of the bias parameters to depend on the values of the cosmological parameters. We found, however, that this dependence is rather mild (the estimates change by at most 4% if Ω_{m} and σ_8 are varied within reasonable intervals). In addition, these computations are done using the Born approximation. Neglecting of higher order corrections mostly affects the bias factor (by a few percent for realistic lens samples), whereas for the correlation factor the errors tend to cancel out. The most important systematic effect on small angular scales is due to the fact that the angular correlation function and the galaxy-galaxy lensing signal cannot be measured below a certain angular scale. This leads to biased estimates of the aperture statistics and thus results in the under- and overestimation of the projected bias and correlation factors, respectively. This is most severe for small lens samples with little angular clustering.

Despite these potential sources of error, we found that for lens galaxies with $r < 20$, the projected bias parameters can be measured with an accuracy of a few percent. Dividing the lenses into a red ($u-r > 2.2$) and a blue sample ($u-r \leq 2.2$), the situation begins to deteriorate, in particular on small scales and for the blue sample. Finally, we investigated the detectability of the time evolution of the projected bias parameters by further splitting the lens samples into low- ($z \leq 0.2$) and high-redshift parts ($z > 0.2$). Only the projected bias factors show significant evolution on scales where the systematic effects of the method are small. The bias factor decreases (increases) with z for the red (blue) galaxies. We found that this evolution is clearly detectable using a lensing survey of $\approx 130 \ \mathrm{deg}^2$.

A An alternative derivation of the ray-tracing formalism

In the case of equal spacings Δw between the lens planes and flat geometry, it is possible to derive the recursion formula for the light ray positions on the lens planes (Eq. 5.12) in a way that makes particularly clear that ray-tracing amounts to solving a discretized version of the equation for the comoving separation vector of two light rays (Eq. 3.57)

$$\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}w^2} = -\frac{2}{c^2}\left[\boldsymbol{\nabla}_\perp\Phi[\mathbf{x}(\boldsymbol{\theta},w),w] - \boldsymbol{\nabla}_\perp\Phi^{(0)}(w)\right] \; . \tag{A.1}$$

We approximate the second derivative on the l.h.s by finite differencing. We form the first derivative of $\boldsymbol{x}$ as follows:

$$\left.\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}w}\right|_{w^{(k)}} \approx \frac{\boldsymbol{x}(w_{\mathrm{U}}^{(k)}) - \boldsymbol{x}(w_{\mathrm{L}}^{(k)})}{\Delta w} \; , \tag{A.2}$$

where Δw is the spacing between two lens planes, and $w_{\mathrm{U}}^{(k)}$ and $w_{\mathrm{L}}^{(k)}$ are the upper and lower boundaries of the slice that is projected onto the k-th plane, respectively. For the second derivative, we find using $w^{(k-1)}$, $w^{(k)}$ and $w^{(k+1)}$ as sampling points:

$$\begin{aligned}
\left.\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}w^2}\right|_{w^{(k)}} &\approx \frac{\boldsymbol{x}(w^{(k+1)}) - \boldsymbol{x}(w^{(k)})}{\Delta w^2} - \frac{\boldsymbol{x}(w^{(k)}) - \boldsymbol{x}(w^{(k-1)})}{\Delta w^2} \\
&= \frac{\boldsymbol{x}(w^{(k+1)}) - 2\boldsymbol{x}(w^{(k)}) + \boldsymbol{x}(w^{(k-1)})}{\Delta w^2} \; .
\end{aligned} \tag{A.3}$$

We discretize the r.h.s. of Eq. (A.1) by replacing it by its mean value in the interval $[w_{\mathrm{L}}^{(k)}, w_{\mathrm{U}}^{(k)}]$:

$$\bar{\boldsymbol{\alpha}}^{(k)}(\boldsymbol{\beta}^{(k)}) = \frac{2}{c^2}\frac{1}{\Delta w}\int_{w_{\mathrm{L}}^{(k)}}^{w_{\mathrm{U}}^{(k)}}\mathrm{d}w \; \boldsymbol{\nabla}_\perp\Phi[\mathbf{x}(\boldsymbol{\theta},w),w] \; , \tag{A.4}$$

where $\boldsymbol{\beta}^{(k)} = \mathbf{x}(\boldsymbol{\theta},w)/w^{(k)}$. We then assume that all light deflections in the k-th slice occur at $w = w^{(k)}$, with a deflection angle $\boldsymbol{\alpha}^{(k)} = \Delta w\,\bar{\boldsymbol{\alpha}}^{(k)}$. Putting these results together, the discrete form of Eq. (A.1) is given by

$$\boldsymbol{x}(w^{(k+1)}) = +2\boldsymbol{x}(w^{(k)}) - \Delta w\,\boldsymbol{\alpha}^{(k)}(\boldsymbol{\beta}^{(k)}) - \boldsymbol{x}(w^{(k-1)}) \; , \tag{A.5}$$

or in angular coordinates,

$$\boldsymbol{\beta}^{(k+1)} = +2\frac{f_K^{(k)}}{f_K^{(k+1)}}\boldsymbol{\beta}^{(k)} - \frac{\Delta w}{f_K^{(k+1)}}\,\boldsymbol{\alpha}^{(k)}(\boldsymbol{\beta}^{(k)}) - \frac{f_K^{(k-1)}}{f_K^{(k+1)}}\boldsymbol{\beta}^{(k-1)} \; . \tag{A.6}$$

This is just Eq. (5.12) for the special case $f_K^{(k-1,k)} = f_K^{(k-2,k-1)} = \Delta w$.

B The minimum area of periodicity of lattice planes

If $\boldsymbol{p} = (p_1, p_2, p_3)^{\mathrm{t}}$ and $\boldsymbol{q} = (q_1, q_2, q_3)^{\mathrm{t}}$ are lattice vectors with $p_i, q_i \in \mathbb{Z}\ \forall\, i$, then $\boldsymbol{p}$ and $\boldsymbol{q}$ span a parallelogram $P_{\boldsymbol{pq}}$ that is periodic along the directions of $\boldsymbol{p}$ and $\boldsymbol{q}$ with a length of periodicity of $|\boldsymbol{p}|$ and $|\boldsymbol{q}|$, respectively. The total area of the parallelogram constructed from $\boldsymbol{p}$ and $\boldsymbol{q}$ is $|\boldsymbol{p} \times \boldsymbol{q}| = |\boldsymbol{n}|$. Here, we show that there is no set of vectors spanning the same plane, but yielding a smaller area of periodicity, if the integer coefficients n_1, n_2, and n_3 are coprime.

To see this, we assume that the components of $\boldsymbol{n} = \boldsymbol{p} \times \boldsymbol{q}$ are coprime. Now let $\boldsymbol{r}$ and $\boldsymbol{s}$ be two lattice vectors lying in the same plane spanning a parallelogram $P_{\boldsymbol{rs}}$ with smaller area than the parallelogram spanned by $\boldsymbol{p}$ and $\boldsymbol{q}$. We can write

$$\boldsymbol{r} = \alpha_1 \boldsymbol{p} + \alpha_2 \boldsymbol{q} \tag{B.1}$$
$$\boldsymbol{s} = \beta_1 \boldsymbol{p} + \beta_2 \boldsymbol{q}\ , \tag{B.2}$$

where $\alpha_i, \beta_i \in \mathbb{Q}$. The area of $P_{\boldsymbol{rs}}$ is given by the modulus of $\boldsymbol{n}' = (n'_1, n'_2, n'_3)^{\mathrm{t}} = \boldsymbol{r} \times \boldsymbol{s}\ (n_i \in \mathbb{Z})$, for which we find

$$\boldsymbol{n}' = (\alpha_1 \beta_2 - \beta_1 \alpha_2)\boldsymbol{n} \equiv A\boldsymbol{n}\ ,\ A \in \mathbb{Q}\ . \tag{B.3}$$

Since we have assumed that $P_{\boldsymbol{rs}}$ has a smaller area than $P_{\boldsymbol{pq}}$, we have $A < 1$, so that we can write $A = a/b$, with $a, b \in \mathbb{Z}$, $b > a$ and a and b coprime. For Eq. (B.3) to hold, b must be a common divisor of n_1, n_2 and n_3, which contradicts our assumptions. Therefore, there can not be a parallelogram with smaller area of periodicity than $P_{\boldsymbol{pq}}$.

C Projection Pursuit Density Estimation

In order to have an independent check of the ICA-based likelihood estimation algorithm described in Chapter 6, we employ the method of *projection pursuit density estimation* (PPDE; Friedman et al. 1984). Like our ICA method, PPDE aims to estimate the joint probability density $p(\boldsymbol{x})$ of a random vector $\boldsymbol{x}$, given a set of observations of $\boldsymbol{x}$. As starting point, an initial model $p_0(\boldsymbol{x})$ for the multidimensional probability distribution $p(\boldsymbol{x})$ has to be provided, for which a reasonable choice is e.g. a multivariate Gaussian with a covariance matrix estimated from the data. The method then identifies the direction $\boldsymbol{\theta}_1$ along which the marginalized model distribution differs most from the marginalized density of the data points and corrects for the discrepancy along the direction $\boldsymbol{\theta}_1$ by multiplying p_0 with a correction factor. This yields a refined density estimate $p_1(\boldsymbol{x})$, which can be further improved by iteratively applying the outlined procedure.

More formally, the PPDE density estimate is of the form

$$p_M(\boldsymbol{x}) = p_0(\boldsymbol{x}) \prod_{m=1}^{M} f_m(\boldsymbol{\theta}_m \cdot \boldsymbol{x}) \,, \tag{C.1}$$

where p_M is the estimate after M iterations of the procedure and p_0 is the initial model. The univariate functions f_m are multiplicative corrections to the initial model along the directions $\boldsymbol{\theta}_m$. The density estimate can be obtained iteratively using the relation $p_M(\boldsymbol{x}) = p_{M-1}(\boldsymbol{x}) \, f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x})$. At the M-th step of the iteration, a direction $\boldsymbol{\theta}_M$ and a function f_M are chosen to minimize the K-L divergence (Eq. 6.49) between the actual data density $p(\boldsymbol{x})$ and the density estimate $p_M(\boldsymbol{x})$,

$$D_{\mathrm{KL}}[p, p_M] = \int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{p_M(\boldsymbol{x})} \,, \tag{C.2}$$

as a goodness-of-fit measure. Only the cross term

$$W(\boldsymbol{\theta}_M, f_M) = -\int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \log p_M(\boldsymbol{x}) \tag{C.3}$$

of the K-L divergence is relevant for the minimization, all other terms do not depend on $\boldsymbol{\theta}_M$ and f_M. By using Eq. (C.1), one sees that the minimum of W is attained at the same location as the minimum of

$$w(\boldsymbol{\theta}_M, f_M) = -\int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \log f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x}) \,, \tag{C.4}$$

which is the expectation value of $\log f_M$ with respect to $p(\boldsymbol{x})$. The data density $p(\boldsymbol{x})$ is unknown; however, the data comprise a set of N samples from this distribution. The expectation value of $\log f_M$ can therefore be estimated by

$$\hat{w}(\boldsymbol{\theta}_M, f_M) = -\frac{1}{N} \sum_{i=1}^{N} \log f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x}_i) \,. \tag{C.5}$$

For fixed $\boldsymbol{\theta}_M$, the minimum of Eq. (C.4) is attained for

$$f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x}) = \frac{p^{\boldsymbol{\theta}_M}(\boldsymbol{\theta}_M \cdot \boldsymbol{x})}{p_{M-1}^{\boldsymbol{\theta}_M}(\boldsymbol{\theta}_M \cdot \boldsymbol{x})} \ , \tag{C.6}$$

where $p^{\boldsymbol{\theta}_M}$ and $p_{M-1}^{\boldsymbol{\theta}_M}$ are the marginal densities of the data and of model density from the $(M-1)$-st iteration along the direction $\boldsymbol{\theta}_M$, respectively. With this, the iterative process that leads to estimates of $\boldsymbol{\theta}_M$ and f_M schematically consists of:

- choosing a direction $\boldsymbol{\theta}_M$,

- computing the marginal densities $p^{\boldsymbol{\theta}_M}$ and $p_{M-1}^{\boldsymbol{\theta}_M}$,

- computing $f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x})$ according to Eq. (C.6),

- computing $\hat{w}(\boldsymbol{\theta}_M, f_M)$

- choosing a new $\boldsymbol{\theta}_M$ that decreases $\hat{w}$

- continuing from step 2 until a convergence criterion is fulfilled.

To efficiently compute the marginals $p^{\boldsymbol{\theta}_M}$ and $p_{M-1}^{\boldsymbol{\theta}_M}$, Monte Carlo samples of these densities are used. Note that the data already comprise a sample of $p(\boldsymbol{x})$; a sample of $p_{M-1}^{\boldsymbol{\theta}_M}$ can be obtained efficiently by an iterative method: since p_{M-1} is similar to p_{M-2}, a subset of the sample from p_{M-1} can be obtained by rejection sampling from the sample from the $(M-2)$-nd step. The remaining data vectors are then drawn by rejection sampling from p_0. For more technical details of the estimation procedure, we refer the reader to Friedman et al. (1984).

Note that the PPDE technique, although using very similar methodology as our ICA-based procedure, is different in the important point that it does not rely on the assumption that a linear transformation of the data leads to statistical independence of the components of the transformed data vectors. It therefore comprises a good test of the validity of this approximation.

D Fisher matrix of the ICA likelihood

In this appendix, we give the derivation of Eq. (6.64). In the general case, the Fisher matrix is given by (e.g. Kendall et al. 1987)

$$
\mathsf{F}_{\alpha\beta} = \left\langle \frac{\partial \log L}{\partial \pi_\alpha} \frac{\partial \log L}{\partial \pi_\beta} \right\rangle \ . \tag{D.1}
$$

In our case, the likelihood depends on cosmological parameters only through the difference between data and model vector, i.e. $\boldsymbol{s} = \breve{\boldsymbol{\xi}} - \breve{\boldsymbol{m}}$ (see Eqs. 6.43 and 6.44). This allows us to write

$$
\frac{\partial \log L(\boldsymbol{s}(\boldsymbol{\pi}))}{\partial \pi_\alpha} = \frac{\partial \log L(\boldsymbol{s})}{\partial s_i} \frac{\partial s_i}{\partial \pi_\alpha} \tag{D.2}
$$

$$
= \frac{\mathrm{d} \log p_{s_i}(s_i)}{\mathrm{d} s_i} \frac{\partial s_i}{\partial \pi_\alpha} \ , \tag{D.3}
$$

where in the last step we have made use of the fact that the likelihood factorizes in the ICA basis. The expression for the Fisher matrix then can be written as

$$
\mathsf{F}_{\alpha\beta} = \sum_{i,j} \left\langle \frac{\mathrm{d} \log p_{s_i}(s_i)}{\mathrm{d} s_i} \frac{\mathrm{d} \log p_{s_j}(s_j)}{\mathrm{d} s_j} \right\rangle \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_j}{\partial \pi_\beta} \tag{D.4}
$$

To compute the expectation value in this equation, we have to average over the likelihood function: $\langle \ldots \rangle = \int \mathrm{d}\boldsymbol{s} \, L(\boldsymbol{s}) \ldots$ We obtain

$$
\mathsf{F}_{\alpha\beta} = \sum_{i \neq j} \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_j}{\partial \pi_\beta} \int \mathrm{d}s_i \frac{\mathrm{d}p_{s_i}(s_i)}{\mathrm{d}s_i} \int \mathrm{d}s_j \frac{\mathrm{d}p_{s_j}(s_j)}{\mathrm{d}s_j} \tag{D.5}
$$

$$
+ \sum_i \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_i}{\partial \pi_\beta} \int \mathrm{d}s_i \, p_{s_i}(s_i) \left(\frac{\mathrm{d} \log p_{s_i}(s_i)}{\mathrm{d}s_i} \right)^2 \ . \tag{D.6}
$$

The integrals in the first term of the right-hand side vanish since the p_{s_i} drop to zero for very large and small values of s_i. This leaves us with

$$
\mathsf{F}_{\alpha\beta} = \sum_i \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_i}{\partial \pi_\beta} \int \mathrm{d}s_i \, p_{s_i}(s_i) \left(\frac{\mathrm{d} \log p_{s_i}(s_i)}{\mathrm{d}s_i} \right)^2 \ . \tag{D.7}
$$

The derivatives in Eq. (D.7) can be strongly affected by noise in the estimated $p_{s_i}(s_i)$, in particular in the tails of the distributions. For their numerical computation, we therefore choose the following four-point finite difference operator (Abramowitz & Stegun 1964):

$$
\frac{\mathrm{d}p}{\mathrm{d}s} = \frac{p(s - 2h) - 8p(s - h) + 8p(s + h) - p(s + 2h))}{12h} + \mathrm{O}(h^5) \ , \tag{D.8}
$$

which we find to be more stable against this problem than its more commonly used two-point counterpart. Because of this potential difficulty, we cross-check our results with the alternative

method provided by Eq. (6.63). This method is significantly slower, but numerically simpler. This is because the derivatives of the log-likelihood in Eq. (6.63) are on average computed close to the maximum-likelihood point, where the likelihood estimate is well sampled. Reassuringly, we find excellent agreement between the two methods. Finally, we have investigated the influence of the choice of the Kernel function K in Eq. (6.42), which might affect the computation of the numerical derivatives. Our results prove to be stable against variation of K, provided that we chose a differentiable Kernel function.

Bibliography

Abramowitz, M. & Stegun, I. A. 1964, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th edn. (New York: Dover)

Anderson, T. W. 2003, An introduction to multivariate statistical analysis, 3rd edn. (John Wiley & Sons)

Bacon, D., Réfrégier, A., & van Waerbeke, L. 2000, MNRAS, 318, 625

Baltz, E. A., Marshall, P., & Oguri, M. 2007, astro-ph/0705.0682

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15

Barlow, R. J. 1991, Statistics (John Wiley & Sons)

Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291

Bennett, C. L., Halpern, M., Hinshaw, G., et al. 1997, in Bulletin of the American Astronomical Society, Vol. 29, Bulletin of the American Astronomical Society, 1353

Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, Phys. Rep., 367, 1

Bernstein, G. M. & Jarvis, M. 2002, AJ, 123, 583

Bertschinger, E. & Gelb, J. M. 1991, Computers in Physics, 5, 164

Blandford, R. & Narayan, R. 1986, ApJ, 310, 568

Brainerd, T. G., Blandford, R. D., & Smail, I. 1996, ApJ, 466, 623

Budavári, T., Connolly, A. J., Szalay, A. S., et al. 2003, ApJ, 595, 59

Chiu, K.-C., Liu, Z.-Y., & Xu, L. 2003, in Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, 751–756

Comon, P., Jutten, C., & Hérault, J. 1991, Signal Processing, 24, 11

Cooley, J. W. & Tukey, J. W. 1965, Math. Comput., 19, 297

Cooray, A. & Hu, W. 2001a, ApJ, 554, 56

Cooray, A. & Hu, W. 2001b, ApJ, 554, 56

Cooray, A. & Hu, W. 2002, ApJ, 574, 19

Cooray, A. & Sheth, R. 2002, Phys. Rep., 372, 1

Crittenden, R., Natarajan, P., Pen, U., & Theuns, T. 2002, AJ, 568, 20

Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371

De Lucia, G. & Blaizot, J. 2007, MNRAS, 375, 2

Dehnen, W. 2001, MNRAS, 324, 273

Dekel, A. & Lahav, O. 1999, ApJ, 520, 24

Desjacques, V., Seljak, U., & Iliev, I. T. 2008, astro-ph/0811.2748

Dietrich, J. P., Erben, T., Lamer, G., et al. 2007, A&A, 470, 821

Dodelson, S. 2003, Modern Cosmology (Amsterdam: Academic Press)

Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2008, astro-ph/0803.0586

Efron, B. & Tibshirani, R. J. 1993, An Introduction to the Bootstrap (Chapman & Hall / CRC)

Eifler, T., Schneider, P., & Hartlap, J. 2008, astro-ph/0810.4254

Einstein, A. 1916, Annalen der Physik, 49, 749

Eisenstein, D. J. & Hu, W. 1999, ApJ, 511, 5

Fisher, R. 1935, J. Roy. Stat. Soc., 98, 39

Friedman, J., Stuetzle, W., & Schroeder, A. 1984, Journal of the American Statistical Association, 79, 599

Frigo, M. & Johnson, S. G. 1998, Proc. ICASSP, 3, 1381

Gelman, A., Carlin, J. B., Stern, H., & Rubin, D. B. 2004, Bayesian Data Analysis (Chapman & Hall/CRC)

Giacconi, R., Rosati, P., Tozzi, P., et al. 2001, ApJ, 551, 624

Gnedin, O. Y., Kravtsov, A. V., Klypin, A. A., & Nagai, D. 2004, ApJ, 616, 16

Goto, T., Yamauchi, C., Fujita, Y., et al. 2003, MNRAS, 346, 601

Gottlöber, S. & Klypin, A. 2008, astro-ph/0803.4343

Gottlöber, S., Klypin, A., & Kravtsov, A. 1999, in Evolution of Large Scale Structure : From Recombination to Garching, ed. A. J. Banday, R. K. Sheth, & L. N. da Costa, 358

Governato, F., Moore, B., Cen, R., et al. 1997, New Astronomy, 2, 91

Guzik, J. & Seljak, U. 2001, MNRAS, 321, 439

Hartlap, J. 2005, Studying Galaxy-Galaxy-Lensing Using Ray-Tracing Simulations, Diploma Thesis (Universität Bonn)

Hastie, T., Tibshirani, R., & Friedman, J. 2001, The Elements of Statistical Learning (Springer)

Hawkins, E., Maddox, S., Cole, S., et al. 2003, MNRAS, 346, 78

Hetterscheidt, M., Simon, P., Schirmer, M., et al. 2005, astro-ph/0606571

Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2008a, astro-ph/0809.5035

Hilbert, S., Metcalf, R. B., & White, S. D. M. 2007a, MNRAS, 382, 1494

Hilbert, S., White, S. D. M., Hartlap, J., & Schneider, P. 2007b, MNRAS, 927

Hilbert, S., White, S. D. M., Hartlap, J., & Schneider, P. 2008b, MNRAS, 386, 1845

Hinshaw, G., Weiland, J. L., Hill, R. S., et al. 2008, astro-ph/0803.0732

Hirata, C. M. & Seljak, U. 2003, Phys. Rev. D, 68, 083002

Hockney, R. W. & Eastwood, J. W. 1988, Computer simulation using particles (Bristol: Hilger)

Hoekstra, H., Mellier, Y., van Waerbeke, L., et al. 2006, ApJ, 647, 116

Hoekstra, H., van Waerbeke, L., Gladders, M. D., Mellier, Y., & Yee, H. K. C. 2002, ApJ, 577, 604

Hoekstra, H., Yee, H. K. C., & Gladders, M. D. 2001, ApJ, 558, L11

Hu, W. & Keeton, C. R. 2002, Phys. Rev. D, 66, 063506

Hyvärinen, A., Karhunen, J., & Oja, E. 2001, Independent Component Analysis (Wiley Interscience)

Hyvärinen, A. & Oja, E. 1997, Neural Computation, 9(7), 1438

Hyvärinen, A. & Oja, E. 2000, Neural Networks, 13(4-5), 411

Jain, B., Seljak, U., & White, S. 2000, AJ, 530, 547

Jaynes, E. 1983, in E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics, ed. R. Rosenkrantz (Kluwer), 151

Jaynes, E. 2003, Probability theory: the Logic of Science (Cambridge University Press)

Jenkins, A., Frenk, C. S., Pearce, F. R., et al. 1998, ApJ, 499, 20

Joachimi, B., Schneider, P., & Eifler, T. 2008, A&A, 477, 43

Johnston, D. E., Sheldon, E. S., Wechsler, R. H., et al. 2007, astro-ph/0709.1159

Jutten, C. & Hérault, J. 1991, Signal Processing, 24, 1

Kaiser, N. 1992, ApJ, 388, 272

Kaiser, N. & Pan-STARRS Collaboration. 2005, in Bulletin of the American Astronomical Society, Vol. 37, 465

Kaiser, N. & Squires, G. 1993, ApJ, 404, 441

Kaiser, N., Wilson, G., & Luppino, G. 2000, astro-ph/0003338

Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 303, 188

Kendall, M. G., Stuart, A., & Ord, J. K. 1987, Kendall's advanced theory of statistics (New York: Oxford University Press, Inc.)

Kilbinger, M., Schneider, P., & Eifler, T. 2006, A&A, 457, 15

Kullback, S. & Leibler, R. A. 1951, Annals of Mathematical Statistics, 22, 79

Landy, S. D. & Szalay, A. S. 1993, ApJ, 412, 64

Lemson, G. & The Virgo Consortium. 2006, astro-ph/0608019

Liang, J.-J., Fang, K.-T., Hickernell, F. J., & Li, R. 2001, Math. Comput., 70, 337

Limber, D. N. 1953, ApJ, 117, 134

Limousin, M., Kneib, J.-P., & Natarajan, P. 2005, MNRAS, 356, 309

Linde, A. 2005, New Astronomy Review, 49, 35

Loredo, T. 1989, in Maximum-Entropy and Bayesian Methods, ed. P. Fougere (Dartmouth: Kluwer Academic Publishers), 81–142

Loverde, M., Hui, L., & Gaztañaga, E. 2008, Phys. Rev. D, 77, 023512

Lynds, R. & Petrosian, V. 1986, in Bulletin of the American Astronomical Society, Vol. 18, 1014

MacKay, D. 2003, Information Theory, Inference and Learning Algorithms (Cambridge University Press)

Mandelbaum, R., Seljak, U., & Hirata, C. M. 2008, Journal of Cosmology and Astro-Particle Physics, 8, 6

Massey, R., Heymans, C., Bergé, J., et al. 2007a, MNRAS, 376, 13

Massey, R., Rhodes, J., Leauthaud, A., et al. 2007b, ApJS, 172, 239

Moessner, R., Jain, B., & Villumsen, J. V. 1998, MNRAS, 294, 291

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493

Neto, A. F., Gao, L., Bett, P., et al. 2007, MNRAS, 381, 1450

Pan, J. & Szapudi, I. 2005, MNRAS, 362, 1363

Peacock, J. 1999, Cosmological Physics (Cambridge University Press)

Peacock, J. A. & Dodds, S. J. 1996, MNRAS, 280, L19

Percival, W. J., Nichol, R. C., Eisenstein, D. J., et al. 2007, ApJ, 657, 645

Pielorz, J. 2008, PhD thesis, Universität Bonn

Press, W. et al. 1992, Numerical Recipes in C (Cambridge University Press)

R Development Core Team. 2007, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0

Refregier, A. & Bacon, D. 2003, MNRAS, 338, 48

Reiprich, T. H. & Böhringer, H. 2002, ApJ, 567, 716

Robert, C. P. & Casella, G. 2005, Monte Carlo Statistical Methods (Springer Texts in Statistics) (Springer)

Robertson, H. P. 1935, ApJ, 82, 284

Rudd, D. H., Zentner, A. R., & Kravtsov, A. V. 2008, ApJ, 672, 19

Schneider, P. 1998, ApJ, 498, 43

Schneider, P. 2005, in Gravitational Lensing: Strong, Weak & Micro, ed. C. Kochanek, P. Schneider, & J. Wambsganss (Springer-Verlag, Berlin)

Schneider, P. 2006a, Extragalactic Astronomy and Cosmology (Berlin: Springer)

Schneider, P. 2006b, in Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, ed. G. Meylan, P. Jetzer, P. North, P. Schneider, C. S. Kochanek, & J. Wambsganss, 1–89

Schneider, P. 2006c, in Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, ed. G. Meylan, P. Jetzer, P. North, P. Schneider, C. S. Kochanek, & J. Wambsganss, 269–451

Schneider, P., Ehlers, J., & Falco, E. E. 1992, Gravitational Lenses (Springer-Verlag)

Schneider, P. & Kilbinger, M. 2007, A&A, 462, 841

Schneider, P. & Rix, H.-W. 1997, ApJ, 474, 25

Schneider, P. & Seitz, C. 1995, A&A, 294, 411

Schneider, P., van Waerbeke, L., Jain, B., & Kruse, G. 1998, MNRAS, 296, 873

Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. 2002a, A&A, 396, 1

Schneider, P., van Waerbeke, L., & Mellier, Y. 2002b, A&A, 389, 729

Schrabback, T., Erben, T., Simon, P., et al. 2007, A&A, 468, 823

Scott, D. W. 1992, Multivariate Density Estimation: Theory, Practice, and Visualization (New York: John Wiley & Sons)

Seitz, S., Schneider, P., & Ehlers, J. 1994, Classical and Quantum Gravity, 11, 2345

Seljak, U. 2000, MNRAS, 318, 203

Seljak, U. & Zaldarriaga, M. 1996, ApJ, 469, 437

Semboloni, E., Mellier, Y., van Waerbeke, L., et al. 2006, A&A, 452, 51

Semboloni, E., van Waerbeke, L., Heymans, C., et al. 2007, MNRAS, 375, L6

Shapiro, C. & Cooray, A. 2006, Journal of Cosmology and Astro-Particle Physics, 3, 7

Sheldon, E. S., Johnston, D. E., Frieman, J. A., et al. 2004, AJ, 127, 2544

Simon, P. 2005, PhD thesis, Universität Bonn

Simon, P., Hetterscheidt, M., Schirmer, M., et al. 2007, A&A, 461, 861

Simon, P., King, L. J., & Schneider, P. 2004, A&A, 417, 873

Smail, I., Hogg, D. W., Yan, L., & Cohen, J. G. 1995, ApJ, 449, L105

Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1991, ApJ, 371, L1

Springel, V. 2005, MNRAS, 364, 1105

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726

Stewart, K. R., Bullock, J. S., Wechsler, R. H., Maller, A. H., & Zentner, A. R. 2008, ApJ, 683, 597

Takada, M. & Jain, B. 2008, astro-ph/0810.4170

Tasitsiomi, A., Kravtsov, A. V., Wechsler, R. H., & Primack, J. R. 2004, ApJ, 614, 533

Tegmark, M. & Peebles, P. J. E. 1998, ApJ, 500, L79

Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22

Tyson, J. A., Wenk, R. A., & Valdes, F. 1990, ApJ, 349, L1

Vale, C. & White, M. 2003, AJ, 592, 699

van Waerbeke, L. 1998, A&A, 334, 1

van Waerbeke, L. et al. 2000, A&A, 358, 30

Venables, W. & Ripley, B. 2002, Modern Applied Statistics with S (Springer)

Villumsen, J. V. 1995, astro-ph/9512001

Walker, A. G. 1936, Proc. Lond. Math. Soc., 42, 90

Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, Nature, 279, 381

Wambsganss, J., Bode, P., & Ostriker, J. P. 2004, ApJ, 606, L93

Weinberg, S. 1972, Gravitation and Cosmology (John Wiley & Sons)

Weinberg, S. 2008, Cosmology (Oxford University Press)

White, M. & Hu, W. 2000, ApJ, 537, 1

White, S. 1993, in Les Houches Session LX, ed. R. Schaeffer, J. Silk, M. Spiro, & J. Zinn-Justin (Elsevier, Amsterdam)

Wilson, G., Kaiser, N., Luppino, G. A., & Cowie, L. L. 2001, ApJ, 555, 572

Wittman, D. et al. 2000, Nature, 405, 143

Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2006, MNRAS, 373, 1159

Yang, X. H., Mo, H. J., Kauffmann, G., & Chu, Y. Q. 2003, MNRAS, 339, 387

Yoo, J., Tinker, J. L., Weinberg, D. H., et al. 2006, ApJ, 652, 26

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1

Zel'Dovich, Y. B. 1970, A&A, 5, 84