Galaxy alignments from multiple angles
Fortuna, M.C.

1 Introduction

A century ago, in 1920, the astronomy community was engaged in the so-called 'the great debate': did the entire Universe consist of just our Galaxy, the Milky Way, or were the observed nebulae non-local objects, being themselves galaxies similarly to our own?[1]

Three years later, at the Mount Wilson Observatory, where the most powerful telescope of the time, the 100-inch Hooker Telescope, was operating, Edwin Hubble put an end to the controversy. Studying the photographic plates, Hubble was able to measure the distance of M31 (Andromeda), finding that it was about a million light-years away. A few years before, Harlow Shapley had measured the size of the Milky Way, assessing a width of about 300 000 light-years (Shapley 1918). This meant that M31 was outside the Milky Way, establishing a new view of the Universe, and opening the door to extragalactic astrophysics and modern cosmology[2].

In 1990, the Hubble Space Telescope was launched. In 1995 it produced one of the highest impact images of astronomy: the Hubble deep field (Fig. 1.1). This high-quality image contains more than 3000 galaxies in just $\sim 0.19\ \mathrm{deg}^2$, and the light of some of these has travelled for 10 billion years to reach us.

In just one century, our perception of the dimensions of Universe had completely changed.

[1] A transcript of the debate can be found at `https://apod.nasa.gov/diamond_jubilee/1920/cs_nrc.html`. For further material, see for example Shapley (1919); Shapley & Curtis (1921)

[2] Our current best estimates of the size of the Milky Way stellar disk and of M31 distance are, respectively, ~ 100000 light years and ~ 2.54 million light years.

Figure 1.1: The Hubble Ultra Deep Field 2012, an improved version of the Hubble Ultra Deep Field image featuring additional observation time. Credit: NASA, ESA, R. Ellis (Caltech), and the HUDF 2012 Team

1.1 A new view of the Universe

In the following years, the measurements of the distances and the velocities of several distant galaxies allowed Hubble to show that they were moving away from our own, and that their recessional velocities were linearly related to their distances: a relation now known as the Hubble-Lemaître law (Hubble 1929; Hubble & Humason 1931; Smith 1979, for a historical review). This was the experimental proof that the Universe was expanding, as already independently predicted a few years before by Friedmann and Lemaître (Friedmann 1922; Lemaître 1931). Due to the expansion of the Universe, the light of distant galaxies is red-shifted (i.e. it is observed at a longer wavelength) and thus we use the redshift, z, as a measure of distance and time; $z = 0$ corresponds to the present Universe, and it increases as we look back in the past.

Our modern view of the Universe is rooted in those years of great discoveries and on the fundamental theoretical works of Einstein, Lemaître, Robertson, Walker and other theorists who contributed to defining the geometrical description of the Universe we rely on. Our current model of the Universe is based on three fundamental assumptions: that we are not located at any special location in the Universe and that on large scales the Universe is isotropic and homogeneous. Starting from these hypotheses and using the theory of General Relativity, it is possible to build a specific class of metrics (Robertson-Walker metrics) and to derive a set of equations that describes the dynamical evolution of our Universe (Friedmann 1924).

Nowadays, the ΛCDM model is the standard cosmological model of reference, which gets its name from the dominant ingredients the Universe is composed of today: an unknown form of energy, called dark energy, that enters into the dynamical equations of the Universe in the form of a cosmological constant called Λ ($\sim 70\%$)[3], and (invisible) cold dark matter (CDM, $\sim 25\%$), originally inferred by its effect on the dynamics of visible matter (Zwicky 1933; Rubin et al. 1980, among others). The remaining components are the ordinary (baryonic) matter ($\sim 5\%$), followed by photons and neutrinos, which together account for less than $\sim 0.01\%$.

The question of what is the nature of dark matter and what is driving the acceleration parametrised by the dark energy are the most fundamental questions in modern cosmology. These involve theoretical as well as

[3]This is required to justify the observation of the late-time acceleration of the expansion of the Universe (Riess et al. 1998; Perlmutter et al. 1999)

experimental investigations (see for example Arun et al. 2017; Kafedžić-Briga & Džaferović-Mašić 2021, for some recent reviews). At the same time, while remaining agnostic on the intrinsic nature of such components, measuring the parameters of the cosmological model with as much precision and accuracy as possible is the main effort of observational cosmology (see Sect. 1.2.2).

In this Thesis, we focus on some specific techniques to infer the cosmological parameters with the main focus on how to control the systematic effects that would lead to incorrect estimates of these. Our subject of study is the alignments of galaxies. These can be intrinsic to the galaxy (what we call intrinsic alignment, IA, see Sect. 1.4) and provide information on the galaxy-dark matter connection, or apparent: in this case we refer to them as gravitational lensing, which we introduce in Sect. 1.3. Gravitational lensing is a powerful probe to explore the dark content of the Universe, as it uses the apparent distortion of galaxy shapes to infer the amount and the spatial distribution of dark matter, dark energy and the geometry of the Universe itself. We focus on modelling possible contaminants to lensing (Chapter 2 and 4) and to exploit lensing to learn more on IA (Chapter 3).

1.2 The Standard Cosmological Model

One of the main consequences of the expansion of the Universe is that it must have been much smaller in the past: in the hypothesis of mass conservation, also the density of the Universe was significantly higher. Thus, the Universe has experienced different phases across its life-time, and, winding back its history, we can reach a moment where the entire space-time was confined in a singularity: the so called Big Bang[4]. From the Big Bang until now, the Universe has continued to expand and cool down: from an initial stage where it was extremely hot and radiation-dominated, and all the particles were in the form of an opaque plasma, to the formation of the first nuclei in what is called the Big Bang Nucleosynthesis. This was followed by the decoupling of matter and radiation and the formation of the first neutral atoms[5] (Gamow 1946, 1948; Alpher & Herman 1948; Alpher et al. 1948). From here, baryonic matter started to assembly due to its grav-

[4]Strictly speaking, there are possible solutions to the Lemaître-Friedmann equations that would admit an ever expanding Universe, but these are excluded by specific observations and constraints on the cosmological parameters at certain redshifts (Boerner & Ehlers 1988).

[5]It is slightly before this point that the Universe entered the matter-dominated era.

itational attraction, forming the first stars and the first galaxies.

A fundamental evidence of the early stage of the Universe consists of the Cosmic Microwave Background (CMB), a relic radiation at the current temperature of ~ 2.725 K which originated at the moment of photon decoupling. The CMB was first predicted by Alpher & Herman (1948) and later detected by Penzias and Wilson (Penzias & Wilson 1965).

It is important to stress that the early Universe was not homogeneous at small scales: the origin of such anisotropies is still unknown, although the theory of inflation (Guth 1981) is generally accepted as the standard paradigm. Early after the equivalence between matter and radiation, matter started to condensate, as an effect of gravitational attraction. However, only dark matter was able to grow: this, not being affected by the electromagnetic radiation, was free to start its collapse, while the baryonic matter (all the visible matter) continued to interact with the surrounding photons. Once matter and radiation decoupled, these dark matter overdensities acted as seeds for structure formation: having higher density than the surrounding matter distribution, they attracted the surrounding baryonic matter (White & Rees 1978). In the standard model of structure formation, the assembly of matter is still driven by the most abundant and simplest form of matter, the collisionless dark matter. This shows the deep connection between dark and visible matter: dark matter forms the cosmic web, providing the gravitational wells where galaxies form and live. Baryonic matter, however, is able to cool via electromagnetic interactions, and can thus contract further, forming denser object such as stars and galaxies. The process of structure formation starts from small objects, which then merge to form larger and more massive ones, in a bottom-up scenario.

1.2.1 The galaxy-halo connection

The dark matter regions that are dense enough to decouple from the cosmic expansion form bounded objects that we refer to as haloes. The most massive haloes sit at the knots of the cosmic web (clusters) and host many galaxies, while less massive ones populate the filaments, and the galaxies they host are referred to as field galaxies (see Fig. 1.2 for a representation of the cosmic web from a N-body simulation). In general, the more massive is a halo, the more galaxies it hosts. In a simplified picture, the first galaxy to be born in a halo sits at its centre; depending on the characteristics of the proto-galaxy and on the surrounding tidal field, the galaxy can orient its major axis (or its spin axis) accordingly to the quadrupole of the gravitational field (see Joachimi et al. 2015; Kiessling et al. 2015, for a re-

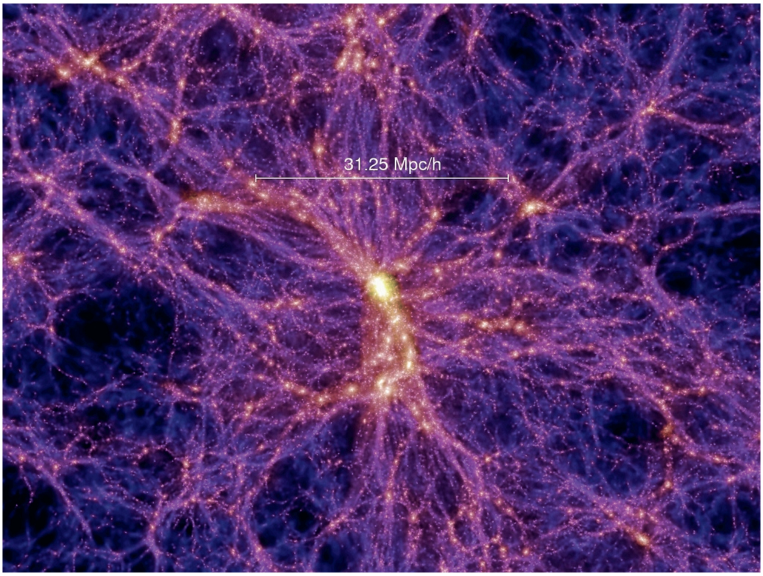

Figure 1.2: A zoom-in on a slice from the Millennium Simulation. A cluster is visible in the centre, surrounded by filaments. Image credit: Springel et al. (2005).

view). Interacting with the surrounding environment, a massive halo keeps attracting smaller structures from the neighbourhood, which become substructures and start orbiting around the halo centre; these sub-haloes can host galaxies themselves: we refer to these galaxies as satellites. Part of these satellites will be tidally stripped, either forming intra-cluster light or accreting onto the central galaxy, which is typically the most massive one in the halo. All of these processes affect the properties of satellites, from their star formation activity to their morphology and their angular momentum, either increasing or destroying the tendency of their major axes to point in the direction of the central galaxy. As the mergers continue, the central galaxy grows, the halo keeps attracting matter from the surrounding filaments, and becomes bigger and bigger: while at large scales the Universe expands, at small scales structures aggregate, contrasting locally the Hubble flow.

The complexity of the relation between galaxies and dark matter makes it impossible to have an exact analytical formalism to describe it. However, their ensemble properties can be captured using a statistical approach. There are several ways to parametrise the galaxy-halo connection: in this Thesis we focus on the halo model and the Halo Occupation Distribution (HOD) (see Cooray & Sheth 2002, for a review). The halo model provides a way to describe the statistical properties of the dark matter haloes, such as their clustering. Dark matter haloes can have different masses: if a halo is massive enough, it can reach virial equilibrium, forming a bounded object, which can then host galaxies. Using N-body simulations, Navarro et al. (1996) showed that the dark matter distribution has a universal profile, with a decreasing density from the core to the outskirts of the halo (NFW profile). An alternative parametrisation which has also shown to fit well the simulations and lensing data is the Einasto profile (e.g. Gao et al. 2008; Mandelbaum et al. 2008). Combining the linear matter density distribution, the distribution of matter within the halo as predicted by the NFW profile and the number density of haloes of a given mass, as predicted by the halo mass function (HMF, Press & Schechter 1974) – now based on N-body simulations (e.g. Jenkins et al. 2001; Lukić et al. 2007; Despali et al. 2016) – it is possible to model the matter density field and predict its 2-point function, the matter power spectrum. The HOD then provides a link between the galaxies and the dark matter haloes, in the form of a probabilistic description of how the galaxies populate the halo. This is based on some observable quantities such as their luminosity and colour, and distinguishes between their type (central/satellites). These models have shown to provide a good fit to a number of observables (e.g. Zheng et al. 2007; Zehavi et al. 2011; More et al. 2011) and can be used to fit cosmological models via combined probes (Cacciato et al. 2013). The halo model is largely used in this Thesis, to link the observed correlation between galaxies and dark matter, for different observables (Chapter 2, 4, 5).

Based on their morphology, galaxies can be broadly divided in two main classes: elliptical and spiral galaxies. Elliptical galaxies are believed to be the result of one or more galaxy mergers, they are characterised by an elliptical shape, old stellar population (emitting more in the red part of the electromagnetic spectrum) and are gravitationally supported by the random motion of their stellar content. Spiral galaxies have a disc-like shape, a young population (emitting more in bluer wavelengths) and are rotationally supported. Due to their typical rest-frame optical wavelength emissions, we will often refer to these two class of objects, respectively, as red and blue galaxies. Cluster galaxies are typically red, as well as the central massive galaxy of big groups. A large fraction of satellite galaxies is also typically red and this correlates with the morphology of the central galaxy (Weinmann et al. 2006). Among the mechanisms responsible for quenching, there are mergers, Active Galactic Nuclei (AGN) and Supernova (SN)

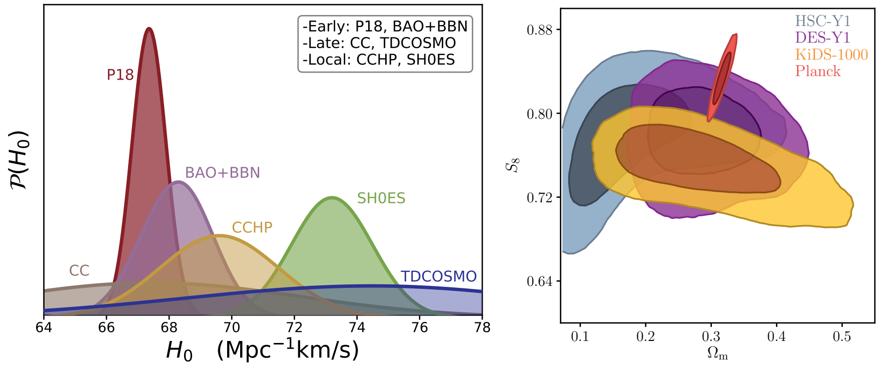

Figure 1.3: *Left:* The posterior distribution of the H_0 parameter as constrained by different analyses. The family of measurements can be broadly divided in three categories: those which rely on measurements of the early physics, such as the CMB (Planck Collaboration et al. 2020) and the Baryonic Acoustic Oscillations (Schöneberg et al. 2019, BAO); those which infer the Hubble constant from the late Universe, using relative galaxy ages (Jimenez & Loeb 2002; Haridasu et al. 2018, Cosmic Chronometers, CC) and the strong lensing of distant quasars (Chen et al. 2019, TDCOSMO); and finally those which measured the value of H_0 in the local universe, via the SNIa, either calibrated using the tip of Red Giant Branch (Freedman et al. 2020, CCHP) or the Cepheids (Riess et al. 2019, SH0EES).Image credit: José Luis Bernal et al. 2020. *Right:* Marginalised constraints for the joint distributions of $S_8 \equiv \sigma_8 \left(\Omega_m/0.3\right)^{0.5}$ and Ω_m obtained by different lensing studies (HSC-Y1, DES-Y1, KiDS-1000) and *Planck*. The contours show the 68% and 95% credible regions. Image credit: Asgari et al. 2021

feedback and ram pressure stripping (Gabor et al. 2010; Zinger et al. 2018, among others).

In this Thesis, we will focus on the dark matter-halo interaction in the late-time Universe, from $z \sim 1.5$ to $z \sim 0.1$, spanning a range of time where the Universe has grown to double its size and has gone from the matter-dominated era to the Λ-dominated one. Our main scientific goal is to use observational data and models of how visible and invisible matter are spatially distributed and oriented to help infer the amount of matter present in the Universe, how it has expanded and clumped over this range of time. This contributes to the main question outlined at the beginning of this introduction, i.e. what are the exact values of the cosmological parameters, to confirm or challenge our current view of the Universe.

1.2.2 The era of precision cosmology and arising tensions

With the advent of space telescopes and the improvement of ground-based facilities, cosmology is facing a new challenge: to constrain the cosmological model to high-precision and accuracy. With the decrease of the statistical error, any systematic effect becomes important and the limits of any model adopted to interpret the data are pushed to their boundaries[6].

In recent years, independent experiments have measured the cosmological parameters with increasing precision. Each experiment is typically able to constrain some parameters better than others, and it is subjected to different systematic effects. The outcomes of these experiments have shown some emerging tension between the different results. The most relevant one involves the expansion rate of the Universe, H_0, as summarised in Fig. 1.3 (left panel). Most notably, the largest tension – 4.2σ – occurs between the value inferred from the analysis of the CMB, and thus of the very early Universe (Planck Collaboration et al. 2020) and a local measure which employs the SNIa as standard candles, calibrated using Cepheids (Riess et al. 2019, SH0ES). These might both point to unknown systematics in the data or new physics, and both directions are current area of research (see Di Valentino et al. 2021, for a recent review on possible solutions to the H_0 tension).

This is, however, not the only tension in the ΛCDM model: *Planck* results are also in mild tension with measurements of the amplitude of matter fluctuations at low redshifts, as constrained by the measurement of the apparent distortion of distant galaxy shapes (e.g. Heymans et al. 2013; Hildebrandt et al. 2017; Troxel et al. 2018; Hikage et al. 2019; Asgari et al. 2021). The light of these galaxies travels across the Universe and gets lensed by the distortion in the space-time due to the presence of matter distribution along the line-of-sight, a phenomenon called weak gravitational lensing (see Sect. 1.3). This can be used to infer the amount of matter in the Universe and by performing a tomographic slicing along the redshift direction to study the evolution of the amplitude of the large scale structures. These two quantities are summarised, respectively, by the cosmological parameters Ω_m and σ_8[7]. It is important to note that, as for the H_0 tension, the results from different groups provide different levels of tension, some of

[6]There is an important distinction between *precision* and *accuracy*: precision regards the size of the error bars and it is improved by the increase in the amount of data; accuracy reflects our ability to recover the *true* value of a parameter, i.e. our ability to model possible sources of biases that would shift the best fit value. Often researchers model our ignorance on systematic effects by adding free parameters that can potentially absorb the bias at the expense of the precision of the constrains: these are called nuisance parameters and the process is called marginalisation.

[7]The parameter σ_8 is defined as the amplitude of the (linear) matter power spectrum, parametrised by the root mean square fluctuations in spheres with a radius of $8h^{-1}\mathrm{Mpc}$ and it has important implications on the growth of fluctuations in the early Universe.

which are compatible with the *Planck* result, as summarised in the right panel of Fig. 1.3.

1.3 Weak Gravitational Lensing

Weak lensing is an extremely powerful probe to investigate the dark content and the geometry of the Universe. Most importantly, it is an orthogonal probe to galaxy clustering. The combined analysis of galaxy clustering and galaxy lensing is the primary goal of current and future cosmological surveys (see also Sect. 1.5).

In light of upcoming surveys which will measure the cosmological parameters with a precision below a per-cent, it is fundamental to understand the possible source of systematics and identify the best approaches to mitigate them. This is also crucial in light of the current tension with *Planck*, as seen in the previous section, in order to ensure a robust estimation of the cosmological parameters and to advocate for the need of an extension to the current cosmological model.

1.3.1 Fundamentals of Gravitational Lensing

Let us first briefly revisit the fundamentals of weak gravitational lensing. General Relativity predicts that a mass distribution distorts the local geometry of the space-time. We know from Fermat's principle that a light ray finds the fastest route in space-time to connect two points: in the absence of any deformation, in a Euclidean geometry, the path followed by the light ray will be a straight line. However, if light is travelling in the proximity of a mass distribution, it will follow a curved trajectory due the distortion of the space-time itself caused by the presence of the mass. In analogy with the distortion in the light-ray trajectory due to the refractive index of a glass lens, we call this process *gravitational lensing*: the mass distribution is the *lens*, while the object from which the light ray is emitted is called *source*. For a review on gravitational lensing, we refer to Bartelmann & Schneider (2001). Typically, the sources are distant galaxies, located behind the lens along the line-of-sight, i.e. $z_{\mathrm{s}} > z_{\mathrm{l}}$, with z_{s} and z_{l} the redshift of the source and the lens, respectively. Since in the majority of cases the distortion is very small[8], we study the induced distortions on an ensemble of source-

[8]This is not the case for strong gravitational lensing, a specific configuration in space that generates a very strong lensing effect. This is a rare phenomenon compared to weak lensing. It is also a promising way to infer cosmological parameters (e.g. Bartelmann &

lens pairs, in a statistical way. Although there is a family of applications to weak lensing (CMB lensing, lensing by clusters, ...), in here we will focus only on two main cases: the lensing effect generated by massive galaxies, which we refer to as galaxy-galaxy lensing (GGL) and the one of the matter distribution along the line-of-sight, which is called cosmic shear. Typically, GGL is used to infer the properties of the class of galaxies used as lenses, such as their total mass and their mass profile (i.e. including the dark matter component), while cosmic shear provides information on the overall matter distribution, growth of structures and the varying of the dark energy parameter, and it can thus be used to infer cosmological parameters, as discussed in the Sect. 1.2.2.

Gravitational lensing acts on the entire image of the background source: it distorts the ensemble of light rays coming from the distant galaxy, which in turn are perceived by the observer as the final (distorted) image of the galaxy. The effect of gravitational lensing is twofold: because the sources are extended objects, the differential deflection of light distorts the images tangentially around the centre of the lens (shearing), and at the same time it magnifies their observed flux due to the local stretch of the space-time, which dilates the image without changing its surface brightness (this is a consequence of the fact that during the process photons are neither absorbed nor emitted). We remind the reader that here we are focusing on weak lensing and thus all of these distortion are small ($\sim 1\%$ of the original light profile).

In the limit of weak gravitational field, the field equations of General Relativity can be linearised: this means that we can treat the deflection induced by an ensemble of point masses as the linear sum of the individual deflections. This provides the framework to treat the deflection caused by an extended mass distribution, which we can treat as the sum of infinitesimal mass elements $\mathrm{d}m$ of volume $\mathrm{d}V$ and volume density $\rho(\mathbf{r})$. Without entering in the details of the derivation, we report here the final expression of the deflection angle $\hat{\alpha}(\xi)$ generated by a mass distribution as sketched in Fig. 1.4, with $\xi = (\xi_1, \xi_2)$ the impact parameter (see e.g. Bartelmann & Schneider 2001, for the full derivation):

$$\hat{\alpha} = \frac{4G}{c^2} \int \mathrm{d}^2\xi' \int \mathrm{d}r_3' \rho(\xi_1', \xi_2', r_3') \frac{\xi - \xi'}{|\xi - \xi'|} \tag{1.1}$$

$$= \frac{4G}{c^2} \int \mathrm{d}^2\xi' \Sigma(\xi') \frac{\xi - \xi'}{|\xi - \xi'|} \,, \tag{1.2}$$

Schneider Che), as discussed in Sect. 1.2.2

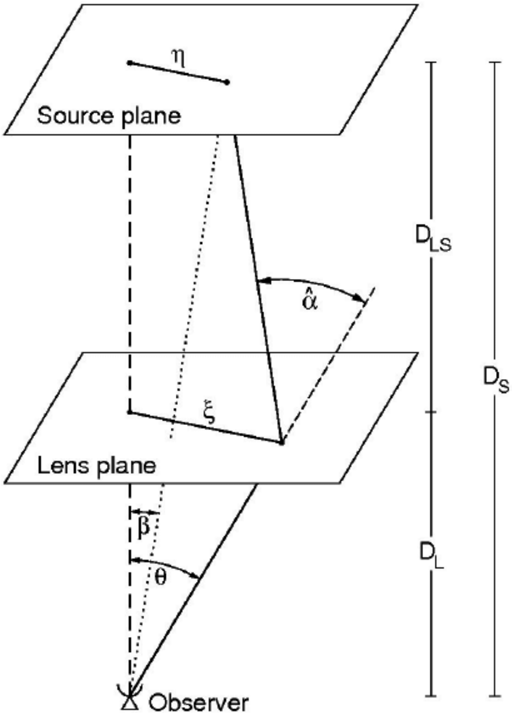

Figure 1.4: Sketch of a typical lensing system. Image credit: Bartelmann & Schneider (2001)

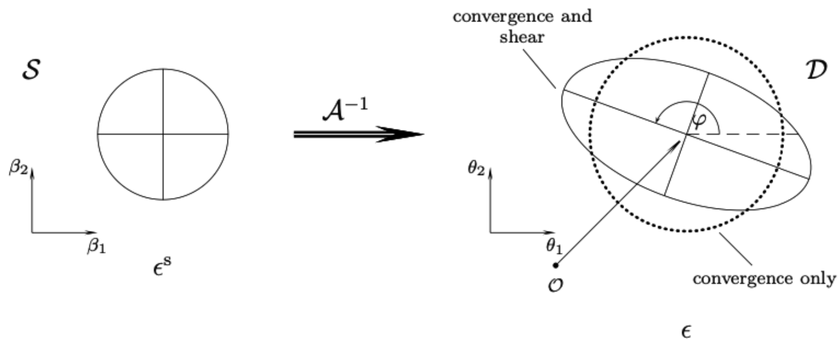

Figure 1.5: A circular source is transformed by the inverse of the Jacobian matrix $\mathcal{A}$: the convergence dilates the image, while the shear changes the axes ratio. Image credit: M. Bradac, from Schneider (2005)

where we have introduced the surface mass density $\Sigma(\xi) \equiv \int \mathrm{d}r_3 \rho(\xi_1, \xi_2, r_3)$, defined as the mass density projected on the plane perpendicular to the line-of-sight.

Let us indicate as β the true angular position of a source galaxy at angular distance D_s from the observer, lensed by a mass distribution located at angular distance D_ds from the source (D_d from the observer). The source will be observed at a new angular position θ given by the vectorial sum of the original position β and the scaled deflection angle $\alpha(\theta)$, as described by the *lens equation*:

$$\beta = \theta - \frac{D_\mathrm{ds}}{D_\mathrm{s}} \hat{\alpha}(D_\mathrm{d}\theta) \equiv \theta - \alpha(\theta) \ . \tag{1.3}$$

To get the final distorted image of an extended object such a galaxy, one should solve the lens equation for each light ray of the image. At first order, a convenient way to visualise the mapping between the original and the observed image is in terms of the Jacobian matrix $\mathcal{A}$, which is the derivative of the original position β with respect to the lensed position θ:

$$\begin{aligned} \mathcal{A} = \frac{\partial \beta}{\partial \theta} = \mathbb{I} - \frac{\partial \alpha}{\partial \theta} &= \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} \\ &= (1-\kappa)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - |\gamma| \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix} \ . \end{aligned} \tag{1.4}$$

Here κ is the convengence, which encodes the isotropic focusing and thus quantifies the changes in size of the image; $\gamma = \gamma_1 + i\gamma_2 = |\gamma| \exp(i2\phi)$ is the complex shear: it quantifies the anisotropic focusing, which yields the distortion in the image. The shear is a spin-2 quantity, i.e. it is invariant under a rotation of 180 deg, as evident from the factor 2 in the phase. From eq. 1.4 follows that a circle would transform into an ellipse, with minor and major axes given by the inverse of the eigenvalues of A. The eigenvectors of $\mathcal{A}$ provide instead the orientation of the ellipse. An illustration of the effect of convergence and shear in a galaxy image is shown in Fig. 1.5. The magnification is the inverse of the determinant of $\mathcal{A}$, $\mu = \frac{1}{\det \mathcal{A}}$.

1.3.2 Cosmic shear

Cosmic shear is the study of the weak cosmological lensing, i.e. of the continuous deflection of a light ray due to the matter inohomegeneities along the line-of-sight. It allows to probe the large scale structures (LSS) by studying the projected correlation of galaxy shapes. As we have seen in the previous section, the presence of matter along the line-of-sight tidally distorts the apparent shapes of the background galaxies and the final image that we receive is the result of the multiple deflections occurred along the way from the emitter to the observer. The formalism to describe cosmic shear has many analogies with the one presented in the previous section and, for the scope of this introduction, we consider it sufficient: the interested reader can refer to e.g. Kilbinger (2015) for a review.

Galaxies are typically non-round objects: we can indicate their intrinsic ellipticity as $\epsilon_{\rm S}$. This is modified by cosmic shear via the introduction of what is called *reduced shear*, g, which is a function of the shear and the convergence acting on the image, $g \equiv \gamma/(1-\kappa)$. In the weak lensing limit, however, since both $|\gamma|$ and κ are $\ll 1$, this simplifies and we can simply consider the distortion as purely due to the shear. The observed ellipticity, ϵ, is thus

$$\epsilon = \frac{\epsilon_{\rm S} + g}{1 + \epsilon_{\rm S} g^*} \approx \epsilon_{\rm S} + \gamma \ . \tag{1.5}$$

The observed ellipticity can be measured from galaxy images with dedicated algorithms and needs to be corrected for the effect of atmospheric blurring and noise. There are several techniques to do this (e.g. by using high fidelity image simulations to calibrate the bias in the recover shape or by using a meta-calibration approach, see for example Mandelbaum (2018) for a review). The observed ellipticity is measured with respect to a reference axis and, as the shear, it can be expressed as a complex quantity: $\epsilon = \epsilon_1 + i\epsilon_2 = |\epsilon|\mathrm{e}^{i2\phi}$ (see Fig. 1.6 for an illustration of the total ellipticity as a function of its two components). In this notation, the absolute value of

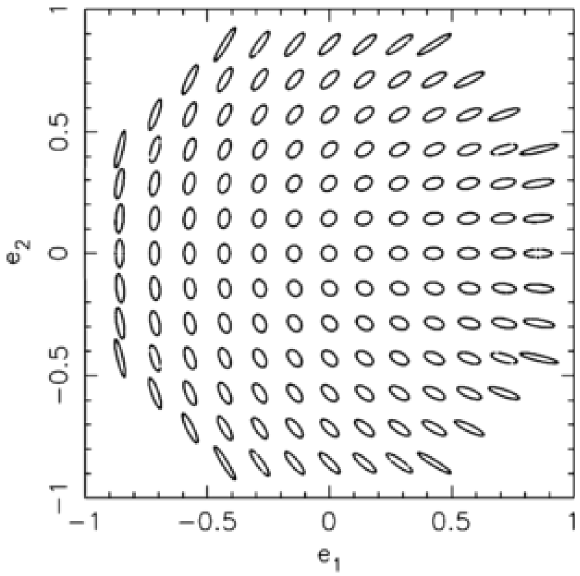

Figure 1.6: The effect of the shear distortions on a circular galaxy, as parametrised by the e_1 and e_2 components. Image credit: Bridle et al. (2009).

the ellipticity is defined as $|\epsilon| = (a-b)/(a+b)$, with a, b the major and minor axes of the ellipse, respectively. In the lack of preferential directions, the ensemble average of the intrinsic shapes would be $\langle \epsilon_s \rangle = 0$ and the observed ellipticity would be an unbiased estimator of the shear. This assumption is, however, violated by any intrinsic alignment (IA), i.e. by the tendency of galaxies to exhibit preferential directions in the sky, as a consequence of their interactions with the LSS. Galaxies form and live embedded in dark matter haloes whose tidal field imprints a preferred alignment to their major axis, as discussed in Sect. 1.2.1. These alignments, being sourced by the LSS, are correlated and are an important contaminant to cosmic shear. IA is the main focus of this work and it is the topic of the next Section.

The efficiency of the lensing depends on the distance between the galaxy and the lens, via a quantity called lensing efficiency:

$$q(\chi) = \int_{\chi}^{\chi_{\rm hor}} \mathrm{d}\chi' n(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}, \tag{1.6}$$

where χ is the comoving distance, $\chi_{\rm hor}$ is the horizon distance, i.e. χ evaluated at infinite redshift; $n(\chi)$ is the source galaxy probability distribution and f_K is the comoving angular distance, which for a flat universe is simply $f_K(\chi) = \chi$. This shows that the lensing efficiency is a function of the ratio $D_{\mathrm{ls}}/D_{\mathrm{s}}$ and the source galaxy distribution $n(\chi)\mathrm{d}\chi = n(z)\mathrm{d}z$. The lensing efficiency is broad and most sensitive to the mass distributions located half-way between the observer and the source.

Because the lensing kernel is a function of comoving distance, it is convenient to split the galaxy sample into tomographic redshift bins, which provide lensing measurements with different weights. By measuring the correlation of galaxy shapes in different z-bins, it is possible to probe the evolution of the Universe at different times. However, to be competitive and to increase the signal-to-noise ratio (S/N), cosmic shear needs to measure galaxy shapes for very wide areas (Amara & Réfrégier 2007), making it impossible to get spectra for all galaxies. Moreover, because the luminosity function increases towards the faint end, the most abundant source objects are typically faint blue galaxies: measuring the spectra of these galaxies is extremely time consuming, and thus different techniques have been developed over time to estimate their redshifts via multi-band photometry. These exploit the redshift-colour relations and are calibrated over subsamples of galaxies for which spectroscopic redshifts are available (see e.g. Wright et al. 2020; Myles et al. 2020, for some recent applications). Weak lensing is mainly sensitive to the mean and the width of the bin redshift distribution (Amara & Réfrégier 2007), however other systematics can depend on other properties (see for example Appendix E in Chapter 2, where we show the impact of the catastrophic outliers on the IA signal), although this is a second order effect, which is not a concern for current surveys but might become important for future ones.

The photo-z provide the 3D spatial distribution of the galaxies, which, at the same time, indicate the time when the light has been emitted. By simultaneously fitting the signal for the different tomographic bins, cosmic shear can probe the growth of structures. However, even if the weights are different for different redshift tomographic bins, these are highly correlated due to the broad kernel of the lensing efficiency and the fact that part of the LSS is in common to all of the bins. Moreover, the uncertainty in the photometric redshifts introduces an overlap between different bins, due to the galaxies that are incorrectly assigned. These factors in practise limit the number of bins one can efficiently obtain, which is typically around five or six in a redshift baseline of $z < 1.2$.

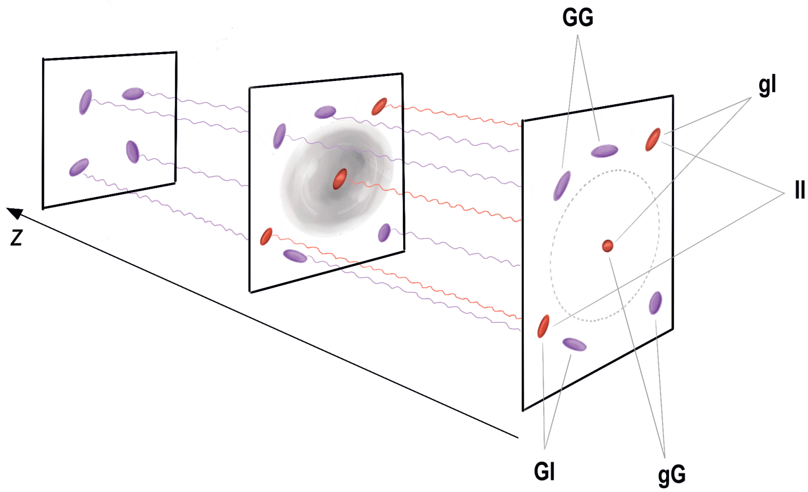

Figure 1.7: A schematic picture of the different terms that arise when correlating galaxy shapes (GG, II, GI), and when correlating galaxy positions and galaxy shapes (gI, gG). Starting from the panel on the left: the light of distant galaxies travels along the line-of-sight and is lensed by a matter overdensity at intermediate redshift (middle panel, indigo ellipses) which align the apparent shapes tangentially to the matter distribution. The galaxies that forms close to the same overdensity are instead intrinsic aligned to the major axis of the halo (red ellipses in the middle panel). The light of all of these galaxies is collected at $z = 0$ where we observe the projected shapes and correlate them to measure either shape-shape correlations, $\langle \epsilon\epsilon \rangle$, or shapes-number density correlations, $\langle \epsilon n \rangle$. For the purpose of illustrating the position-shape correlations, the central galaxy is shown as a circular point in the rightmost panel.

1.4 Intrinsic Alignment

As mentioned in the previous section, IA is the tendency of galaxies to have an intrinsic orientation imprinted at the moment of galaxy formation and evolved over time due to the continuous interaction with the surrounding tidal fields (Catelan et al. 2001). We have seen in Sec. 1.2.1 that galaxies form inside dark matter haloes and inherit from them their angular momenta and orientations via complex mechanisms that happen at the level of the proto-galaxy (see Joachimi et al. 2015; Kiessling et al. 2015; Kirk et al. 2015, for dedicated reviews). Our knowledge of these mechanisms is still limited and we mainly rely on hydrodinamical simulations to get insights. We can broadly distinguish two main regimes: the alignment of central galaxies, which dominates at large scales, and the one of satellite galaxies, which becomes important at small scales (Schneider & Bridle 2010). These two have an intrinsically different nature due to their different formation history and their location within the halo. Moreover, a clear distinction between the alignment of blue, disc-like galaxies and red, elliptical galaxies has also emerged both in observations (Hirata et al. 2007; Joachimi et al. 2011; Mandelbaum et al. 2011; Singh et al. 2015; Johnston et al. 2019) and simulations (Chisari et al. 2015b; Velliscig et al. 2015b; Tenneti et al. 2016), with the former showing no alignment within current precision, while the latter have a clear alignment signature, which further depends on secondary galaxy parameters, such as their luminosity, as we will discuss later.

In terms of its role as contaminant to weak lensing, we identify two IA correlations that contribute to the final observed measured in cosmic shear (see Fig. 1.7): the alignment between intrinsic galaxy shapes (intrinsic-intrinsic term, II), $\langle \epsilon_{\rm s} \epsilon_{\rm s} \rangle$, which is sourced by the tendency of galaxies that form close to a same overdensity to share the same alignment, and the correlation between the intrinsic shape of a galaxy an the shear of another galaxy, $\langle \gamma \epsilon_{\rm s} \rangle$. The latter is called the gravitational-intrinsic term (GI), and it is caused by a matter overdensity that simultaneously shears a background galaxy and aligns a foreground one, close to the mass distribution itself. It is the predominant contamination since it acts between galaxies that are separated in redshift (cross-terms in the projected correlation), while the II is only relevant when galaxies have similar redshifts, as they need to be physically connected to generate such term. The observed correlation is the sum of all of the terms:

$$\langle \epsilon \epsilon \rangle = \langle \gamma \gamma \rangle + \langle \epsilon_{\rm s} \epsilon_{\rm s} \rangle + \langle \gamma \epsilon_{\rm s} \rangle \ . \tag{1.7}$$

Here, $\langle\gamma\gamma\rangle$ is the term sourced by cosmic shear: it is clear that in order to recover the correct cosmic shear signal from the observed correlation we need to model the other two terms and account for them in the cosmological parameter inference.

Similarly, when measuring the galaxy position - shape correlation, as in the galaxy-galaxy lensing, we have

$$\langle\epsilon n\rangle = \langle\gamma n\rangle + \langle\epsilon_{\rm S} n\rangle \; , \tag{1.8}$$

i.e. we need to account for the correlation between the position of a galaxy and the intrinsic orientation of another, nearby one. In this case, for the IA to be significant, the galaxies need to be physically connected and thus a cut in the lens-source pair separation is typically sufficient to remove the contamination.

Measuring IA poses the same challenges as measuring lensing: we have to deal with noisy, point-spread function (PSF)-convolved data, and we need to remove from the signal the contribution from lensing. To measure galaxy shapes there are different algorithms available, employing different methods (Kaiser et al. 1995; Kaiser, Squires & Broadhurst Mil; Melchior et al. 2011, e.g.). In this Thesis we focus on two methods and in Chapter 3 we compare their effect on the associated IA signal.

To measure the IA signal, it is convenient to look at the projected correlation function between galaxy positions and galaxy orientations, since it has a better S/N compared to the II signal and can be considered a proxy to the GI signal. It is fundamental to have good estimates of the galaxy positions to ensure that the galaxy pair is truly physically connected and minimise the lensing contribution: this is one of the main limitations to IA studies. For this reason, it is in general preferred to work with spectroscopic redshifts. In this thesis we investigate the signal in a photometric sample (KiDS-1000 LRGs, Chapter 3), which requires a full modelling of the possible contaminants (lensing and magnification).

Because the selection function of a survey implicitly selects galaxies with different properties (typically because of cuts in apparent magnitude, but it can also be in colours, or due to the presence of fibre/slits) the average properties of the galaxy sample will differ from survey to survey, and even more importantly will possibly change along the redshift baseline. It is thus of crucial importance to study how IA depends on observable galaxy properties such colour and luminosity, and to try to trace back the underlying physical relation, such as the formation history, the mass of the halo and the orbital time. Most of this Thesis tries to deal with such complexity, by

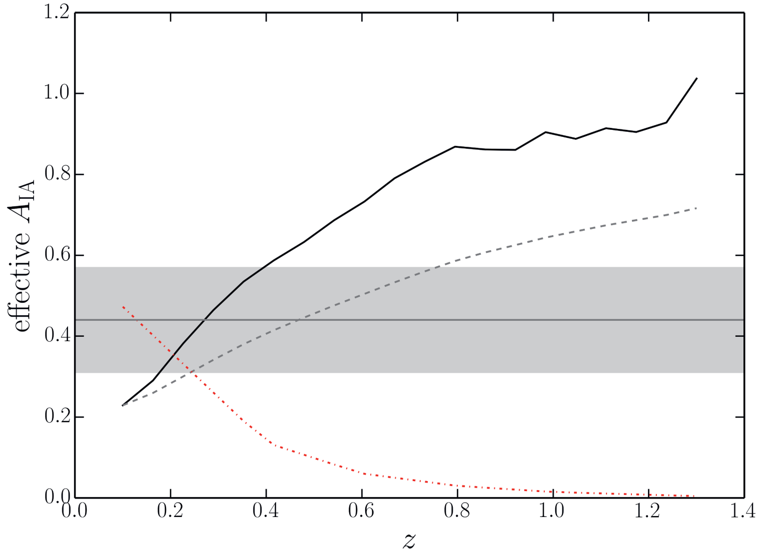

Figure 1.8: The effective IA signal input in the data vector (solid black line) and recovered by the NLA model (gray band). The dashed gray line shows the cumulative mean of the input IA signal, while the red dash-dotted line the ratio of IA over lensing at different redshifts (see the text for more details). Image credit: Robertson et al. (2020)

both informing the models with direct observation of intrinsic alignment of specific galaxy sub-samples (Chapter 3), and trying to connect them to the properties of the halo they reside on (Chapter 4). We dedicate a full analysis of how these differences propagate in the contamination of a lensing survey in Chapter 2.

It is interesting to note that even though a typical lensing survey – which is flux-limited – selects galaxies with a stronger IA signal at high redshifts (IA is typically stronger for luminous red central galaxies), the IA impact on cosmic shear is strongest in the low-redshift bins, where lensing is less efficient (see Sec. 1.3.2). We investigate this effect in Robertson et al. (2020) and present it in Fig. 1.8. Here we compare the best-fit amplitude for a KiDS-like analysis (grey band) from the model presented in (Fortuna et al. 2021a) and the input IA signal in the data vector (see Chapter 2 for more details). The best-fit value is lower than the mean IA signal and broadly corresponds to the signal present only in the low-redshift bins. This can be understood by considering the lensing efficiency over redshift: at high$-z$ the relative importance of IA over lensing decreases significantly, as we can see from the red dot-dashed line, which shows the ratio of the projected angular power spectra $C_{\rm IA}^{\rm (ij)}(\ell)/C_{\rm GG}^{\rm (ij)}(\ell)$ evaluated at the multipole $\ell = 1000$ as a function of the redshift of the foreground bin j. Hence, although the actual IA signal is larger in the high-z tomographic bins, it impacts the best fit parameters less, because an error in the estimated IA signal there has a negligible effect on the inferred $C(\ell)$.

This is an important finding for CMB lensing studies, as it shows that although their results are sensitive to the IA amplitude, they can expect a minor contribution at the redshift probed by their analysis. Similarly, this might suggest that removing the first tomographic bin might be a safer choice for cosmic shear surveys.

1.5 Data

In the context of cosmology, it is common to divide the surveys into categories based on their constraining power. We refer to the ongoing surveys as Stage-III and the next generation surveys as Stage-IV. Currently there are three ongoing weak lensing surveys: the Kilo Degree Survey (KiDS, Kuijken et al. 2019), the Dark Energy Survey (DES, Abbott et al. 2021) and the Hyper Supreme-Cam Survey (HSC, Aihara et al. 2018). In this Thesis we focus on KiDS, which we use both as a reference to simulate a generic Stage-III survey on Chapter 2, and to investigate in depth the properties of one sub-sample of galaxies, the LRGs, in Chapter 3 and 4. We measure the shapes of these galaxies to study its IA signal using a moment-based algorithm, from which we measure the shapes as second-moments of the surface brightness. A full characterisation of this sample is then performed in Chapter 3, via the use of a halo model fit.

Stage-IV surveys include both space-based and ground-based surveys, which are designed to be synergic in their observing strategies. These are *Euclid* (Laureijs et al. 2011), the Vera C. Rubin Observatory, previously known as the Large Synoptic Survey Telescope (LSST, Abell et al. 2009) and the Nancy Roman Space Telescope, previously known as Wide Field InfraRed Survey Telescope (WFISRT, Spergel et al. 2015). A substantial part of the cosmology community is currently involved in forecasting the abilities of Stage-IV surveys to constrain the cosmological parameters and on identifying strategies to mitigate the systematic errors. We dedicate two chapters to these analyses, for both a generic Stage-IV survey (Chapter 2) and an LSST-like survey (Chapter 5). An extended version of the pipeline developed for these two projects is currently used by the Euclid-IA working group to forecast the impact of IA in *Euclid* and to provide a self-calibration strategy to mitigate it. The pipeline is specifically designed to interface with CosmoSIS (Zuntz et al. 2015) a software for cosmological parameter fitting. The pipeline is fully modular, such that any ingredient of the model (the concentration-mass relation, the halo mass function etc...) can be replaced by the user with minimal editing. The pipeline allows to compute all the 2-point functions in Fourier space which enter into the clustering, IA and their cross-correlation statistics.

1.6 This Thesis

The main goal of this Thesis is to investigate the IA of galaxies to mitigate its impact on cosmic shear. Based on previous observations that IA depends on galaxy properties such as galaxy colour, type (central/satellite) and luminosity, in **Chapter 2** we study the impact of not accounting for the variety of IA signatures in the data when analysing cosmic shear signal and assessing the level of bias that would arise for Stage-III and Stage-IV surveys. We also provide an analytical prescription of how it should be modelled to fully account for such complexity and identify a minimal model that would at least capture the variation of the IA contamination across the tomographic redshift bins, as a consequence of the evolution on the composition of the galaxy population in a flux-limited survey. We find that the IA contamination is largest at low redshift, where lensing is less efficient and that the behaviour of the luminosity dependence at faint luminosities is crucial to assess the level of contamination. That was, however, the regime where less data were available and thus we provide a double scenario forecast, extrapolating over the most extreme regimes allowed by the data.

In **Chapter 3** we use the high quality data from KiDS and we focus on exploring the IA dependence luminosity and redshift for a highly S/N sample with precise photometric redshifts, the luminous red galaxies (LRGs). This sample is ideal to cover part of the unconstrained range of the luminosity dependence, and the high S/N allows us to completely disentangle the luminosity and the redshift dependence of the IA signal with specific cuts. We found the data to favour a broken power-law scenario for the luminosity dependence with a knee at $L \lesssim 3.2{\times}10^{10} h^{-2} L_{\odot}$ in the r-band; they also do not exhibit any redshift dependence from $z = 0.2$ to $z = 0.8$.

In **Chapter 4** we extend the investigation of the galaxy sample used in Chapter 3 to constraint IA and we explore the dark matter properties of these galaxies to provide a more direct link between the halo mass and the IA signal. We use GGL to infer the mass profile of the LRGs and to assess the fraction of satellites in the sample. This is also crucial information to properly model the IA signal, since it is observed that centrals and satellites exhibit different IA signals (Johnston et al. 2019). To link the visible and invisible component of the galaxy and infer the galaxy properties we make use of the halo model, with the halo occupation distribution based on the conditional luminosity function (Yang et al. 2003). We also measure the IA signal sourced by the background galaxies which are physically connected to the LRGs, selecting only the sources for which $|z_\mathrm{l} - z_\mathrm{s}| < 0.2$.

In **Chapter 5** we focus on the other main effect of lensing which is often neglected: magnification. We consider the hypothesis of improving the cosmological constraints by including the effect of magnification and provide a forecast for the LSST. We include the large scale IA in the model and make use of the same halo model formalism as in Chapter 4. For this project, I developed a pipeline for the theoretical prediction of the clustering and the luminosity function. This includes the modelling of how the luminosity function is affected by the photometric redshift uncertainty. We find that the improvement in the recovered cosmological parameters by adding the magnification is negligible, while the effect of ignoring magnification can severely bias the cosmological parameters.