# Alma Mater Studiorum · Università di Bologna

**Dipartimento di Fisica e Astronomia "Augusto Righi"**
**Corso di Laurea magistrale in Astrofisica e Cosmologia**

# A new implementation of an optimal filter for the detection of galaxy clusters through weak lensing

Tesi di laurea

Presentata da:
**Gabriele Panebianco**

Relatore:
**Prof. Lauro Moscardini**

Corelatore:
**Prof. Matteo Maturi**

*Dedicato a Salvatore Puglisi, nonno superman e sarto sopraffine, che continua ad essere la mia
Stella Polare...*

## Abstract

We developed a new version of a `C++` code, *Get the Halo* 2021, that implements the optimal linear matched filter presented in Maturi et al. (2005).

Our aim is to detect dark matter haloes of clusters of galaxies through their weak gravitational lensing signatures applying the filter to a catalogue of simulated galaxy ellipticities. The dataset represents typical data that will be available thanks to the *Euclid* mission, thus we are able to forecast the filter performances on weak lensing data obtained by Euclid. The linear matched filter is optimised to maximise the signal-to-noise ratio (S/N) of the detections and minimise the number of spurious detections caused by superposition of large-scale structures; this is achieved by suppressing those spatial frequencies dominated by the large-scale structure contamination.

We compared our detections with the true population of dark matter haloes used to produce the catalogue of ellipticities. We were able to confirm the expectations on the filter performance raised by Maturi et al. (2005) and Pace et al. (2007). We found that $S/N \approx 7$ can be considered as a reliable threshold to detect haloes through weak lensing as $\simeq 83\%$ of our detections with $S/N > 7$ were matched to the haloes; this is consistent with Pace et al. (2007). The purity of our catalogues of detections increases as a function of S/N and reaches 100% at $S/N \approx 10.5 \div 11$. We were also able to confirm that the filter selects preferentially the haloes with redshift between 0.2 and 0.5, that have an intermediate distance between observer and nearby background sources, condition that maximises the lensing effects. The completeness of our catalogues is a steadily growing function of the mass until $(4 \div 5)\, M_\odot\, h^{-1}$, where it reaches values $\approx 58 \div 68\%$.

Our algorithm might be used to enhance the reliability of the detections of the *AMICO* code (Bellagamba et al. 2018), the optimal linear matched filter implemented in the *Euclid* data analysis pipeline to identify galaxy clusters in photometric data (Euclid Collaboration et al. 2019).

***Keywords***— cosmology:theory, dark matter, gravitational lensing

## Abstract

In questo lavoro di tesi abbiamo sviluppato una nuova versione di un codice `C++` che implementa il *filtro adattivo lineare ottimale* presentato da Maturi et al. (2005).

Il nostro obiettivo principale è rilevare aloni di materia oscura di ammassi di galassie dalle loro tracce nei dati di lensing debole, applicando il filtro a un catalogo di ellitticità di galassie simulate. Esso rappresenta dati tipici che saranno disponibili grazie alla missione *Euclid*, e siamo pertanto in grado di effettuare previsioni sulla performance del filtro sui dati di lensing deboli che saranno ottenuti da Euclid. Il filtro è costruito per massimizzare il rapporto segnale-rumore (S/N) delle rilevazioni di aloni e minimizzare il numero di rilevazioni spurie causate dalla sovrapposizione di strutture su grande scala; ciò è effettuato sopprimendo le frequenze spaziali dominate dalla contaminazione della struttura a larga scala.

Abbiamo confrontato le nostre rilevazioni con il reale insieme di aloni di materia oscura usati per produrre il catalogo delle ellitticità. Abbiamo potuto confermare le previsioni sulle prestazioni del filtro di Maturi et al. (2005) e Pace et al. (2007). Abbiamo trovato che un S/N $\approx 7$ può essere considerato come una soglia affidabile per la rilevazione di aloni tramite lensing debole in quanto $\simeq 83\%$ delle nostre rilevazioni con S/N$> 7$ sono state associate ad aloni; questo è in accordo con le previsioni di Pace et al. (2007). La purezza dei nostri cataloghi di rilevazioni aumenta con il S/N e raggiunge il 100% a S/N $\approx 10.5 \div 11$. Abbiamo confermato che il filtro trova aloni preferibilmente con redshift tra 0.2 e 0.5, che hanno una distanza intermedia fra l'osservatore e le sorgenti di background vicine, condizione che massimizza gli effetti di lensing. La completezza dei nostri cataloghi è una funzione crescente della massa fino a $(4 \div 5)\,\mathrm{M}_\odot\,h^{-1}$, dove vale $\approx 58 \div 68\%$.

Il nostro algoritmo potrebbe essere usato per migliorare l'affidabilità delle rilevazioni del codice *AMICO* (Bellagamba et al. 2018), il filtro adattivo lineare ottimale implementato nella pipeline di analisi dati di *Euclid* per identificare ammassi di galassie nei dati fotometrici (Euclid Collaboration et al. 2019).

***Keywords***— cosmologia:teoria, materia oscura, lensing gravitazionale

# Contents

# Introduction

The detection of dark matter haloes is a crucial task in Astrophysics and Cosmology in order to understand how our Universe works. Measuring their number and distribution can put constraints on the cosmological models and the structure formation theory used to describe the Universe and how galaxies and clusters of galaxies were made. Detecting dark matter haloes is also important because it is possible that dark haloes might exist in the Universe, i.e. dark matter haloes of substantial mass that for unknown reasons fail to produce light. Claims have been made in the literature that dark mass concentrations were significantly detected through their weak-lensing signal (Erben et al. 2003). If confirmed, those detections would be extremely interesting because they would show that large, dark haloes could exist while for some reason failing to emit light, either as stellar light from galaxies or X-ray emission from hot intracluster plasma. Since the baryon fraction in clusters should faithfully reflect the cosmic mixture of baryonic and dark mass (Ettori et al. 2006), detections of truly dark cluster-sized haloes could shed doubt on our understanding of the formation of non-linear cosmic structures. The only way to measure their presence is by measuring the gravitational distortion effect they would have on light, i.e. through their gravitational lensing signal.

One of the main sources of error in these measurements is due to the fact that the dark matter distribution in the Universe is not a discrete distribution that contains only the haloes, but it can be thought as a continuous distribution where the haloes are peaks of mass density embedded into a lower density *large-scale structure*, the cosmic web that connects them. When someone measures the inhomogeneities of projected mass distribution through lensing observations the signal of haloes is overlaid by the lensing signal of the large-scale structure in front of and behind the haloes. Being approximately a Gaussian random field, lensing by large-scale structure adds peaks and troughs to the signal which can be mistaken for haloes, so part of the claimed dark-halo detections can actually be peaks in the random weak gravitational lensing signal of the large-scale structure.

It is not possible to strictly separate these two types of signal because of the unsharp boundary between the haloes and the large-scale structure, but it is possible to define a typical scale that helps to discriminate between their lensing signals. The large-scale structure can be considered as composed by dark-matter haloes of a broad and continuous

mass range. At each cosmological epoch there is a characteristic scale length, the *nonlinear scale*, at which the variance of the dark matter density contrast becomes unity. This scale length separates the small-scale regime where the dark matter power spectrum is dominated by the contributions of presumably virialised haloes from the large-scale regime where the dark-matter density can be considered as a linear superposition of linearly evolved perturbation modes.

This scale suggests the operational definition of the lensing signal due to the haloes as the signal contributed by non-linear, gravitationally bound, virialised structures; the lensing signal due to the large scale structure is instead the signal contributed by the linearly evolved matter distribution beyond the non-linear scale. It was first shown in Maturi et al. (2005) and Maturi et al. (2007) that these definitions can be used to construct a linear matched filter capable to detect the weak lensing signal of dark matter haloes of galaxy clusters while also suppressing the signal of the large scale structure, thus reducing the spurious detections attributed to random peaks in the projected mass density of the large scale structure. The underlying assumption of this approach is that the filter is searching for those haloes that do create the non-linear power spectrum, while the linearly evolved structures are treated as a noise component influencing the shape of the matched filter.

This thesis is structured as follows. In chapter 1 we review those aspects of *Cosmology* that are needed to properly understand the rest of the thesis, and the same is done in chapter 2 for the theory of *Gravitational lensing*, with a particular emphasis on the *weak lensing* regime. In chapter 3 we describe the theoretical construction of the optimal linear matched filter to detect dark matter haloes and the program used to implement it, called **Get the Halo 2021** (GtH21). In chapter 4 we apply it to a set of simulated data (a catalogue of galaxy ellipticities) and compare the detections found with the haloes of the simulation. We end this thesis with a final chapter, *Conclusion and future perspectives*, where we review and summarise the main results of this thesis. We also discuss why they are relevant, how they might impact future works and even some of the results of the *Euclid* mission.

# Chapter 1

# An introduction to cosmology

This chapter contains an introduction to the general aspects of the standard model of Cosmology that are relevant for the following chapters. The standard model consists of a modelization for the cosmological background, which is a homogeneous and isotropic solution of the field equations of General Relativity, and a theory for structure formation. The background model is described by the Robertson - Walker metric, in which hypersurfaces of constant time are homogeneous and isotropic three-dimensional spaces, either flat or curved, and change with time according to a scale factor which depends on time only. The dynamics of the scale factor is determined by two equations, Friedmann's equations, which follow from Einstein's field equations given the highly symmetric form of the metric. Current theories of structure formation assume that structures grow via gravitational instability from initial seed perturbations: their origin is yet unclear, but is a possible result of a cosmological inflation. Most common hypotheses lead to the prediction that the statistics of the seed fluctuations is Gaussian. Their amplitude is low for most of their evolution so that linear perturbation theory is sufficient to describe their growth until late stages.

In this chapter we are going to review in section 1.1 how Friedmann cosmological models are built, and we shall introduce the *Lambda*CDM model that describes our Universe (section 1.1.2). In section 1.2 we are going to describe the concept of *density perturbations*, structures that will ultimately grow into the cosmic structures we observe today, such as galaxies and clusters of galaxies, presented in section 1.3. Section 1.4 describes the main quantities of *clustering studies* that will be relevant in the following chapters. The main references for this chapter are Bartelmann et al. (2001), Schneider (2015) and Carroll (2004).

## 1.1  On the Friedmann cosmological models

Modern cosmological models assume gravitation is described by *Einstein's General Theory of Relativity*. General Relativity describes space-time as a four-dimensional manifold whose metric tensor $g_{\mu\nu}$ is considered as a dynamical field. The four spacetime coordinates are generally expressed as $(x^\mu) = (c\,t, \boldsymbol{x}) = \left(c\,t, x^1, x^2, x^3\right)$. Two events in spacetime with coordinates differing by $\mathrm{d}x^\mu$ are separated by $\mathrm{d}s$, with $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\mathrm{d}x^\nu$. The dynamics of the metric is governed by Einstein's field equations, which relate the Einstein tensor to the stress-energy tensor of the matter contained in spacetime (Carroll 2004):

$$\mathcal{G}_{\mu\nu} = \frac{8\pi G}{c^4}\,\mathcal{T}_{\mu\nu} + \Lambda g_{\mu\nu}. \tag{1.1}$$

The quantities that appear in the equations are:

- *Metric tensor $g_{\mu\nu}$.* Symmetric tensor used to compute the distance between two points of the spacetime manifold.

- *Einstein's tensor $\mathcal{G}_{\mu\nu} = \mathcal{R}_{\mu\nu} - \dfrac{1}{2}\mathcal{R}\,g_{\mu\nu}$.* Symmetric tensor, it is a function of the *Ricci-Curbastro curvature tensor* $\mathcal{R}_{\mu\nu}$ that describes the curvature of the spacetime manifold. It also uses its trace, the *Ricci scalar* or *curvature scalar*:

$$\mathcal{R} = \mathcal{R}^\mu{}_\mu = g^{\mu\nu}\mathcal{R}_{\mu\nu}. \tag{1.2}$$

- *Stress-energy-momentum tensor $\mathcal{T}_{\mu\nu}$.* Symmetric tensor that describes the distribution in spacetime of mass and energy, that are the sources of the gravitational field in General Relativity.

- *Universal gravitational constant $G$,* with $G \approx 6.67 \times 10^{-11}\mathrm{N\,m^2\,kg^{-2}}$.

- *Speed of light in vacuum $c$,* with $c \approx 2.99 \times 10^5\,\mathrm{km\,s^{-1}}$.

- *Cosmological constant $\Lambda$.* It describes the "dark energy" component of the Universe that is causing the acceleration of its expansion.

Modern cosmological models are based on the following two principles:

- **Copernican principle:** Earth is not in a privileged position to observe the Universe.

- **Cosmological principle:** The Universe is spatially homogeneous and isotropic.

While these principles are not true on a local scale, several cosmological observations (such as the isotropy of the *Cosmic Microwave Background Radiation*) suggest that they are valid on scales larger than hundreds of Mpc $h^{-1}$ (Schneider 2015). The cosmological principle considerably constrains the admissible form of the metric tensor. Spatial coordinates which are constant for fundamental observers are called *comoving coordinates*. In these coordinates, the mean motion is described by $\mathrm{d}x^i = 0$, and hence $\mathrm{d}s^2 = g_{00}\mathrm{d}t^2$. Greek indexes run over $0\ldots3$ and Latin indexes run over the spatial indexes $1\ldots3$ only. If we require that the proper time of fundamental observers equals the cosmic time, this implies $g_{00} = c^2$. Isotropy requires that clocks can be synchronized such that the space-time components of the metric tensor vanish, $g_{0i} = 0$, otherwise these components would identify one particular direction in spacetime. The spatial metric can only isotropically contract or expand with a *scale factor* $a(t)$ which must be a function of time only, because otherwise the expansion would be different at different places, violating homogeneity. Homogeneity also implies that all quantities describing the matter content of the Universe, e.g. density and pressure, can be functions of time only. The metric can be written as

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\mathrm{d}l^2, \tag{1.3}$$

where $\mathrm{d}l$ is the line element of the three-dimensional space, that must be spherically symmetric because of isotropy. Homogeneity permits us to choose an arbitrary point as coordinate origin. It is possible to prove (Bartelmann et al. 2001) that the cosmological principle constrains the metric into a form known as *Friedmann-Lemaître-Robertson-Walker* (FLRW):

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\left[\frac{1}{1-Kr^2}\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right], \tag{1.4}$$

where $(r, \theta, \phi)$ are the spatial spherical coordinates. We can assume them to be adimensional and the scale factor to have the dimension of a length (but this is not the only prescription that can be used). The constant $K \in \mathbb{R}$ is the *curvature constant*: if $K = 0$ the Universe has a *flat geometry*; if $K > 0$ the Universe is closed with a *spherical geometry*; if $K < 0$ the Universe is open with a *hyperbolic geometry*. One could always redefine the radial coordinate so that the curvature constant can only assume 3 possible integer values according to its sign: $K = 0$ for a flat Universe; $K = +1$ for a closed spherical Universe; $K = -1$ for an open hyperbolic Universe. Unless we state otherwise, we use here the conventional choice according to which $K \in \mathbb{R}$, but we must mention that this is not the only possible choice. The line element can be written as

$$\mathrm{d}l^2 = \mathrm{d}\chi^2 + f_K^2(\chi)\mathrm{d}\Omega^2, \tag{1.5}$$

where $\Omega$ is the solid angle and $\chi$ is a radial coordinate that makes the radial term independent of the curvature. The function $f_K$ is equal to the radial coordinate ($r = f_K(\chi)$), but expressed as a function that homogeneity requires being either a trigonometric, linear,

or hyperbolic function of $\chi$, depending on whether the curvature $K$ is positive, null or negative:

$$r = f_K(\chi) = \begin{cases} K^{-\frac{1}{2}} \sin\left(K^{\frac{1}{2}}\chi\right) & K > 0 \\ \chi & K = 0 \\ (-K)^{-\frac{1}{2}} \sinh\left((-K)^{\frac{1}{2}}\chi\right) & K < 0 \end{cases}. \tag{1.6}$$

The inverse function can be written as:

$$\chi = S_K(r) = \begin{cases} K^{-\frac{1}{2}} \arcsin\left(K^{\frac{1}{2}}r\right) & K > 0 \\ r & K = 0 \\ (-K)^{-\frac{1}{2}} \operatorname{arcsinh}\left((-K)^{\frac{1}{2}}r\right) & K < 0 \end{cases}. \tag{1.7}$$

The FLRW metric is *not static*: the proper distance between two points with fixed spatial coordinates (that differ from $r$) is a function of time and can be computed as

$$d_p(t,r) = \int_0^r \sqrt{g_{11}}\, \mathrm{d}r' = a(t)\, S_K(r) = a(t)\, \chi. \tag{1.8}$$

The current epoch is usually referred with the subscript 0, so $t_0$ is the current time and $a_0$ is the current scale factor. They can be used to define the *comoving distance $d_c$* as the current proper distance; the reference frame of the comoving coordinates has the property of being independent of the time evolution of the scale factor, so two points with fixed spatial coordinates do not change their comoving distance:

$$d_c(r) = d_p(t_0) = a_0\, S_K(r) = a_0\, \chi = \frac{a_0}{a(t)}\, a(t)\, \chi = \frac{a_0}{a(t)}\, d_p(t,r). \tag{1.9}$$

The time derivative of the logarithm of the scale factor is called the *Hubble parameter* and it is the relative expansion rate of the Universe:

$$H(t) \equiv \frac{\mathrm{d}\ln a(t)}{\mathrm{d}t} = \frac{\dot{a}}{a}. \tag{1.10}$$

All measurements confirm that it is a positive quantity, so the Universe is *expanding*; see Schneider (2015) for a discussion about cosmological observations. Due to the expansion of space, photons are redshifted while they propagate from the source to the observer; this effect is known as *cosmological redshift*. The redshift of a photon emitted with wavelength $\lambda_{em}$ and observed with wavelength $\lambda_{obs}$ is defined as:

$$z \equiv \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} = \frac{\Delta\lambda}{\lambda_{em}}. \tag{1.11}$$

The redshift of a photon emitted at a cosmological time $t$ and received at the current time $t_0$ can be expressed as a function of the scale factor by using only the FLRW metric (1.4) (Bartelmann et al. 2001):

$$1 + z = \frac{a_0}{a(t)}. \tag{1.12}$$

Measuring $z > 0$ confirms that $a_0 > a$ and that the Universe is expanding. The redshift is also an observational proxy for the distance in space and time of the sources.

### 1.1.1 On the Friedmann-Lemaître equations

To complete the construction of the Friedmann cosmological models we need a set of equations that would describe how physical quantities of the Universe evolve with time. The content of the Universe is described as made up by uniform "cosmic fluids" that have an energy-momentum tensor given by:

$$\mathcal{T}_{\mu\nu} = -p\,g_{\mu\nu} + \left(p + \rho\,c^2\right)u_\mu u_\nu, \tag{1.13}$$

where $p$ is the fluid pressure, $\rho$ the density, $u_\mu = \dot{x}_\mu$ the time derivative of the comoving coordinates. If we solve Einstein's equations (1.1) with this definition for the energy-momentum tensor and the FLRW metric (1.4) the *Friedmann-Lemaître* equations can be obtained:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} + \frac{\Lambda}{3}, \tag{1.14}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda}{3}. \tag{1.15}$$

Equation (1.14) is also known as *Friedmann's equation*, while equation (1.15) as the *acceleration equation*. They can be combined to yield the *adiabatic equation*:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[a^3(t)\rho(t)c^2\right] + p(t)\frac{\mathrm{d}}{\mathrm{d}t}\left[a^3(t)\right] = 0. \tag{1.16}$$

This can be interpreted as the first law of thermodynamics in the cosmological context: the first term $a^3\rho$ is proportional to the energy contained in a fixed comoving volume, and hence the equation states that the change in "internal" energy equals the pressure times the change in proper volume. It can be also written as the *fluid equation*:

$$\dot{\rho} + 3H(\rho + p) = 0. \tag{1.17}$$

To build a cosmological model we need to express $\rho$, $p$ and $a$ as functions of time; we need a third equation other than equations (1.14) and (1.17). We assume that the *equation of state* that expresses pressure as a function of density is linear:

$$p = w\rho c^2. \tag{1.18}$$

If we insert equation (1.18) into equation (1.17) we solve the fluid equation and obtain:

$$\rho(a) = \rho_0\,a^{-3(1+w)}. \tag{1.19}$$

The constant $w$ is the *state parameter* and describes the properties of the cosmic fluid. There are three main different kinds of cosmic fluids:

- **Matter** or *dust*: they are massive non relativistic particles at rest in the comoving frame. The pressure of matter is thus 0, $w = 0$, $\rho \propto a^{-3}$. This slope tells us that density decreases only because the volume increases.

- **Radiation**: cosmic fluid made up by photons or relativistic particles. Neutrinos can be considered a radiative cosmic fluid if they are assumed to be massless or relativistic. The trace of their energy-momentum tensor must be null, and this constraint yields $p = \frac{1}{3}\rho c^2$, $w = 1/3$ and $\rho \propto a^{-4}$. Density decreases more steeply than matter because not only the volume increases, but the photons also lose energy because of the cosmological redshift.

- **Dark energy**: it is defined as a cosmic fluid with $w < -\frac{1}{3}$. An *ordinary* cosmic fluid made up by matter and radiation has $0 \leq w < 1$ and can only decelerate the expansion of the Universe, as can be seen from equation (1.15). In order to explain the acceleration of the Universe that can be observed today, one needs a cosmic fluid with $w < -\frac{1}{3}$. The Cosmological constant is the most basic form of dark energy; it is a cosmic fluid with $w = -1$ and constant density and pressure.

A cosmological model is also defined by its *cosmological parameters*. One of the most relevant is the *Hubble constant*, that is defined as the value of the Hubble parameter defined in equation (1.10) at the present time

$$H_0 \equiv H(t_0) = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}. \tag{1.20}$$

The value of $H_0$ is still uncertain, so it is common practice to express it through $h$; this is an open problem in Cosmology that we address in the following section 1.1.2. Current estimates are around $h \approx 0.7$ (Schneider 2015). The time scale for the expansion of the Universe is the inverse Hubble constant or *Hubble time* $H_0^{-1} \approx 10^{10}\,\mathrm{yr}\,h^{-1}$. It can be seen as an order-of-magnitude estimate for the age of the Universe.

Another important quantity is the *critical density* of the Universe

$$\rho_c(t) \equiv \frac{3\,H^2(t)}{8\pi G}. \tag{1.21}$$

Its current value is $\rho_{c,0} \approx 1.9 \times 10^{-29}\,h^{\,2}\mathrm{g\,cm^{-3}}$. The density of a cosmic fluid in units of $\rho_c$ is known as the *density parameter* of the cosmic fluid:

$$\Omega(t) = \frac{\rho(t)}{\rho_c(t)} = \frac{8\pi G}{3H^2}\,\rho. \tag{1.22}$$

The density parameter of the cosmological constant is

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H_0^2}. \tag{1.23}$$

Friedmann's equation (1.14) can be written as:

$$(\Omega_{ord} + \Omega_\Lambda) - 1 = \frac{K}{H^2 a^2},\tag{1.24}$$

where $\Omega_{ord}$ is the sum of the density parameters of the ordinary cosmic fluids of the Universe (matter and radiation). The total density parameter of the Universe today is $\Omega_{tot,0} = \Omega_{ord,0} + \Omega_\Lambda$. This is an extremely important cosmological parameter, since it is directly tied to the geometry of the Universe:

- If $\Omega_0 < 1$ then $\rho < \rho_c$ and $K < 0$, the Universe is open, its geometry hyperbolic.

- If $\Omega_0 = 1$ then $\rho = \rho_c$ and $K = 0$, the Universe is flat. The observations today suggest that the Universe is indeed flat. The critical density is then the density necessary to make the Universe flat.

- If $\Omega_0 > 1$ then $\rho > \rho_c$ and $K > 0$, the Universe is closed, its geometry spherical.

Another cosmological parameter is the *deceleration parameter* $q_0$:

$$q_0 \equiv -\frac{a\,\ddot{a}}{\dot{a}^2}.\tag{1.25}$$

Current measures suggest that is it negative, so the Universe is accelerating its expansion.

## 1.1.2   On the ΛCDM model

The standard cosmological model that describes our Universe is also known as ΛCDM model. It contains matter, radiation and dark energy described through the Cosmological constant. The value of the main cosmological parameters, as derived from the most recent analysis of the Cosmic Microwave Background Radiation data of the *Planck mission* (Aghanim et al. 2020), can be found in table 1.1. The Hubble constant is value is yet uncertain, but it is around $h \approx 0.70$. One of the main problems in Cosmology today is the *tension* in the measurements of the Hubble constant: independent measurements yield different and not consistent values. The Planck analysis, based on cosmological data, yields $H_0 = (67.4 \pm 0.5)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Local measurements based on the period-luminosity relation of the Cepheids yield $H_0 = (74 \pm 3)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, giving rise to a discrepancy at the $\sim 2\sigma$ level. Strong lensing measurements that use the time delay of multiple images yield different values up to $H_0 = (79 \pm 4)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. These estimates are discussed in Schneider (2015). This is still an open problem.

Today we have evidence of two kinds of matter: ordinary or *baryonic matter*, made up by the particles of the Standard Model of Particle Physics (such as protons, neutrons and electrons); *dark matter* made up by unidentified yet particles that interact only gravitationally with each other and the baryons. There is a lot of observational and

theoretical evidence for the existence of dark matter, like the rotation curves of neutral gas in spiral galaxies, the mass estimate of clusters of galaxies thanks to optical, X and lensing observations, theory of cosmic structure formation and so on. Studies of the large scale structure of the Universe suggest that the dark matter needed is *Cold Dark Matter*, not relativistic and consistent with a *bottom-up* scenario for cosmic structure formation, where small halos were the first to be born and then merged into bigger halos by hierarchical aggregation. The content of matter in the Universe is $\approx 85\%$ dark matter and $\approx 15\%$ baryonic matter.

The Universe contains two candidates for radiation today, photons and neutrinos (if massless as predicted by the standard model of Particle Physics). The energy density contained in photons today is determined by the temperature of the Cosmic Microwave Background, a black-body radiation with temperature $T_{CMB,0} = 2.73\,\mathrm{K}$. The density parameter of the radiation is $\Omega_{rad,0} = 2.4 \times 10^{-5}\,h^{-2}$. Like photons, neutrinos were produced in thermal equilibrium in the hot early phase of the Universe. Interacting weakly, they decoupled from the cosmic plasma when the temperature of the Universe was $K_B T \approx 1\mathrm{M\,eV}$. When the temperature of the Universe dropped to $K_B T \approx 0.5\,\mathrm{M\,eV}$, electron-positron pairs annihilated to produce gamma rays. The annihilation heated up the photons but not the neutrinos which had decoupled earlier. Hence the neutrino temperature is lower than the photon temperature by an amount determined by entropy conservation. It can be shown $T_{\nu,0} = \left(\dfrac{4}{11}\right)^{\frac{1}{3}} T_{CMB,0} = 1.95\,\mathrm{K}$. The density parameter of one neutrino species (a total of three exist) is $\Omega_{\nu,0} = 2.8 \times 10^{-6}\,h^{-2}$. The radiation energy content of the Universe today is negligible with respect to the other components.

Measurements from the CMBR indicate that the total density parameter of the Universe is consistent with being $\Omega_{tot,0} = 1$. These measurements and the evidence for the acceleration of the Universe led to theorize the existence of a Cosmological constant with $\Omega_{0,\Lambda} \approx 0.7$.

## 1.2   On density perturbations and their evolution

The distribution of galaxies on the sky is not uniform or random, rather they form clusters and groups of galaxies. Also clusters of galaxies are not distributed uniformly, but their positions are *correlated*, grouped together in superclusters. The three-dimensional distribution of galaxies is also known as *large-scale structure*. It is sometimes referred to as the *cosmic web* since matter tends to aggregate into structures made up by massive filaments of matter separated by giant voids; clusters are found at the intersections of these filaments. It proves that the Universe is inhomogeneous on small scales, so Friedmann models are an approximation valid on large scales.

The standard model for the formation of structures in the Universe assumes that there were small fluctuations at some very early initial time, which grew by gravitational

**Table 1.1:** Cosmological parameters from Planck 2018 (Aghanim et al. 2020). We report the density parameters of baryons, dark matter, Hubble constant, age of the Universe, matter density parameter, dark energy density parameter.

| Parameter | Value |
|---|---|
| $\Omega_{0,bar}h^2$ | $0.0224 \pm 0.0001$ |
| $\Omega_{0,DM}h^2$ | $0.120 \pm 0.001$ |
| $H_0[\mathrm{km\,s^{-1}\,Mpc^{-1}}]$ | $67.4 \pm 0.5$ |
| $t_0[\mathrm{Gyr}]$ | $13.801 \pm 0.024$ |
| $\Omega_{0,m}$ | $0.3111 \pm 0.0056$ |
| $\Omega_{0,\Lambda}$ | $0.6889 \pm 0.0056$ |

instability. To be more accurate, those were fluctuations in the metric tensor field in the Universe, that caused fluctuations also in the gravitational potential field and in the density field. Although the origin of the seed fluctuations is yet unclear, they possibly originated from quantum fluctuations in the very early Universe, which were blown up during a later inflationary phase. The fluctuations in this case are uncorrelated and the distribution of their amplitudes is Gaussian. Gravitational instability leads to a growth of the amplitudes of the relative density fluctuations, described by the *density contrast field*

$$\delta(\boldsymbol{x}, t) \equiv \frac{\rho(\boldsymbol{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \tag{1.26}$$

where $\bar{\rho}(t)$ denotes the mean cosmic matter density at time $t$. The CMB anisotropy suggests that at $z \sim 1000$, $|\delta| \ll 1$, while galaxies today have $|\delta| \gg 1$. As long as the relative density contrast of the matter fluctuations is much smaller than unity, they can be considered as small perturbations of the otherwise homogeneous and isotropic background density, and linear perturbation theory suffices for their description. The general consensus is that the density of the Universe was first dominated by radiation in its early times; at an epoch known as the *equivalence* ($a_{eq}$) the contribution of matter and radiation is equal. After the equivalence the density of the Universe was dominated by weakly interacting dark matter (when $a \gg a_{eq}$). Today the density of the Universe has two similar contributions from the matter and the Cosmological constant, which is slightly more dominant, as described in section 1.1.2.

A relevant scale length in the evolution of the perturbations is the *horizon scale*, that is the scale of the causally connected regions in the Universe:

$$\lambda_H(t) = a(t) \int_0^t \frac{c}{a(t')} \, \mathrm{d}t'. \tag{1.27}$$

Its exact value is a function of the cosmological model and parameters, but in general it is inversely proportional to the *Hubble parameter* $\lambda_H \propto cH^{-1}$. A perturbation of (comoving) wavelength $\lambda$ is said to "enter the horizon" when $\lambda = \lambda_H(a)$. This is a relevant scale for structure growth. Perturbation theory shows that linear density fluctuations, i.e. perturbations with $\delta \ll 1$, that have *not* entered the horizon at a cosmic epoch with scale factor $a$ ($\lambda > \lambda_H(a)$) grow like:

$$\delta(a) \propto a^q = \begin{cases} a^2 & \text{if } a < a_{eq} \\ a & \text{if } a > a_{eq} \end{cases} . \tag{1.28}$$

The cosmic microwave background reveals temperature fluctuations of order $\frac{\delta T}{T} \approx 10^{-5}$ on large scales. By the Sachs-Wolfe effect (Sachs et al. 1967), these temperature fluctuations reflect density fluctuations $\delta$ of the same order of magnitude. The cosmic microwave background originated at $a \approx 10^{-3} \gg a_{eq}$, well after the Universe became matter-dominated. Equation (1.28) then implies that the density fluctuations today, expected from the fluctuations $\delta \approx 10^{-5}$ at $a \approx 10^{-3}$, should only reach a level of $10^{-2}$. They should not have been able to enter the non-linear regime at $\delta \sim 1$ and form structures with $\delta \gg 1$ like the galaxies and clusters we see today. This discrepancy is resolved thanks to dark matter. The cosmic microwave background displays fluctuations in the baryonic matter component only. If there is an additional matter component that only couples through weak interactions, fluctuations in that component could grow as soon as it decoupled from the cosmic plasma, well before photons decoupled from baryons to set the cosmic microwave background free. Such fluctuations could therefore easily reach the amplitudes observed today, and thereby resolve the apparent mismatch between the amplitudes of the temperature fluctuations in the cosmic microwave background and the present cosmic structures. Baryons could then form structures more quickly because they would collapse not in a uniform background potential, but in a potential characterized by the existence of dark matter halos (*baryon catch-up*). This is one of the strongest arguments for the existence of a dark matter component in the Universe.

Equation (1.28) describe the growth of perturbations in a flat Universe. It cannot be analytically generalised to curved universes, but numerical simulations show the effect of curvature for $a > a_{eq}$ (it is negligible in the early universe, for $a < a_{eq}$). When $\Omega_{tot,0} < 1$ the gravitational pull of the density fluctuations is weaker and the expansion of the universe is stronger with respect to a flat Universe and perturbations for $a > a_{eq}$ grow as $\delta(a) \propto a^q$ with $q < 1$. When $\Omega_{tot,0} > 1$ the gravitational pull of the density fluctuations is stronger and the expansion of the universe is weaker with respect to a flat Universe and perturbations for $a > a_{eq}$ grow as $\delta(a) \propto a^q$ with $q > 1$.

Equation (1.28) shows that perturbations outside the horizon always grow. When they enter the horizon this is not true in general because gravity is not the only relevant interaction between particles and other physical interactions must be considered too. If $\lambda < \lambda_H(a_{eq})$, the perturbation enters the horizon while radiation is still dominating the

expansion. Until $a_{eq}$, the expansion time-scale, the *Hubble time* $\tau_H = H^{-1}$, is determined by the radiation density $\rho_{rad}$. In this case it is shorter than the collapse time-scale of the dark matter, because $(G\,\rho_{rad})^{-1/2} < (G\,\rho_{DM})^{-1/2}$. The radiation-driven expansion is "faster" than the dark-matter collapse and prevents dark-matter perturbations from collapsing. Since light can only cross regions that are smaller than the horizon, the suppression of growth due to radiation is restricted to scales smaller than the horizon; larger-scale perturbations remain unaffected. This explains why the horizon size at the equivalence epoch $\lambda_H(a_{eq})$ sets an important scale for structure growth.

The epoch $a_{ent}$ at which a density perturbation with comoving wavelength $\lambda$ enters the horizon is found by solving $\lambda = \lambda_H(a_{ent})$. This is:

$$\lambda \propto \begin{cases} a_{ent} & a_{ent} \ll a_{eq} \\ a_{ent}^{1/2} & a_{eq} \ll a_{ent} \ll a_0 \end{cases}. \qquad (1.29)$$

It is convenient to decompose the density contrast $\delta$ into Fourier modes, because in linear perturbation theory individual Fourier components evolve independently. The Fourier transform of $\delta(\boldsymbol{x})$ is:

$$\tilde{\delta}(\boldsymbol{k}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \delta(\boldsymbol{x})\, e^{-i\,\boldsymbol{k}\cdot\boldsymbol{x}}\, \mathrm{d}^3\boldsymbol{k}. \qquad (1.30)$$

Let us note that the dimensions of $\tilde{\delta}(\boldsymbol{k})$ are the ones of a volume and that, since $\delta \in \mathbb{R}$ we must have $\tilde{\delta}(\boldsymbol{k})^* = \tilde{\delta}(-\boldsymbol{k})$. Let us also introduce the *Dirac's delta distribution* and its effect when applied to a generic function $f$:

$$\delta_D^{(3)}(\boldsymbol{k}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} e^{i\,\boldsymbol{k}\cdot\boldsymbol{x}}\, \mathrm{d}^3\boldsymbol{x}, \qquad (1.31)$$

$$f(\boldsymbol{x}) = \int_{\mathbb{R}^3} f(\boldsymbol{k})\, \delta_D^{(3)}(\boldsymbol{x} - \boldsymbol{k})\, \mathrm{d}^3\boldsymbol{k}. \qquad (1.32)$$

The assumed Gaussian density fluctuations $\delta(\boldsymbol{x})$ at the comoving position $\boldsymbol{x}$ can completely be characterized by their *Power Spectrum* $\mathcal{P}(k)$, that will be properly defined later in section 1.4:

$$\langle \tilde{\delta}(\boldsymbol{k})\, \tilde{\delta}^*(\boldsymbol{k}') \rangle = (2\pi)^3\, \mathcal{P}(k)\, \delta_D^{(3)}(\boldsymbol{k} - \boldsymbol{k}'), \qquad (1.33)$$

where the asterisk denotes complex conjugation. Strictly speaking, the Fourier decomposition is valid only in flat space, but it can be applied here because at early times space is flat in any cosmological model, and at late times the interesting scales $\lambda = \frac{2\pi}{k}$ of the density perturbations are much smaller than the curvature radius of the Universe. The power spectrum is proportional to the average square amplitude of a perturbation with wave vector $k$, $\mathcal{P}(k) \propto \langle |\delta(\boldsymbol{k})|^2 \rangle$. Averaging in the Fourier space means averaging every wave vector $\boldsymbol{k}$ whose wave number is $|\boldsymbol{k}| = k$.

Let us consider now the primordial perturbation spectrum at some very early time, i.e. when it was formed at the end of the inflation: $\mathcal{P}_i(k) \propto |\delta(\boldsymbol{k})|^2$. Since the density contrast

grows as in equation (1.28), the spectrum grows as $\mathcal{P}(k;a) \propto a^{2q}$. At $a_{ent}$, equation (1.29) and $k \propto \lambda^{-1}$ show that the spectrum has changed to

$$\mathcal{P}(k;a_{ent}) \propto a_{ent}^{2q}\, \mathcal{P}_i(k) \propto \begin{cases} k^{-4}\, \mathcal{P}_i(k) & a_{ent}(k) < a_{eq} \\ \mathcal{P}_i(k) & a_{ent}(k) > a_{eq} \end{cases} , \qquad (1.34)$$

where $a_{ent}(k) < a_{eq}$ describes the scales that enter the horizon before the equivalence (and thus have their growth suppressed), and $a_{ent}(k) > a_{eq}$ are the large scales (small $k$) not reached by the horizon before the equivalence.

It is commonly assumed that the total power of the density fluctuations at $a_{ent}$ should be scale-invariant. This implies $k^3\,\mathcal{P}(k) = const$ or $\mathcal{P}(k) \propto k^{-3}$. Accordingly, the primordial spectrum has to scale with $k$ as $\mathcal{P}_i(k) \propto k$. This scale-invariant spectrum is called the *Harrison-Zel'dovich spectrum*. We can rewrite (1.34) as

$$\mathcal{P}(k;a_{ent}) \propto \begin{cases} k^{-3} & a_{ent}(k) < a_{eq} \\ k & a_{ent}(k) > a_{eq} \end{cases} . \qquad (1.35)$$

An additional complication arises when the dark matter consists of particles moving with a velocity comparable to the speed of light. In order to keep them gravitationally bound, density perturbations then have to have a certain minimum mass, or equivalently a certain minimum size. All perturbations smaller than that size are damped away by free streaming of particles. Consequently, the density perturbation spectrum of such particles has an exponential cut-off at large k. This clarifies the distinction between hot and cold dark matter: *Hot dark matter* (HDM) consists of fast particles that damp away small-scale perturbations, while *Cold dark matter* (CDM) particles are slow enough to cause no significant damping.

The normalization of the power spectrum is not fixed by theoretical models and must be measured from observations. One way to do it is to measure the normalization of the anisotropies of the cosmic microwave background radiation (CMBR). Another is by the local variance of galaxy counts: galaxies are supposed to be biased tracers of underlying dark-matter fluctuations. By measuring the local variance of galaxy counts within certain volumes, and assuming an expression for the bias, the amplitude of dark-matter fluctuations can be inferred. Conventionally, the variance of galaxy counts $\sigma_{8,gal}$ is measured within spheres of radius $8\,h^{-1}$ Mpc, and the result is $\sigma_{8,gal} \approx 1$ at the present time. The corresponding variance $\sigma_8$ of matter density fluctuations can be found by assuming a bias mechanism that allows to pass from galaxy counts to matter density; the bias is a prediction of galaxy formation theories still unknown and under debate. A third way to get the normalization is by the local abundance of galaxy clusters: they form by gravitational instability from dark-matter density perturbations. Their spatial number density reflects the amplitude of appropriate dark-matter fluctuations in a very sensitive manner. It is therefore possible to determine the amplitude of the power spectrum by demanding that the local spatial number density of galaxy clusters be reproduced.

At late stages of the evolution and on small scales, the growth of density fluctuations begins to depart from the linear regime. When density fluctuations grow in a non-linear way, fluctuations of different size interact and the evolution of $\mathcal{P}(k)$ becomes so complicated it needs to be evaluated numerically. The non-linear evolution of the density fluctuations is crucial for accurately calculating weak lensing effects by large-scale structures. Non-linearly evolved density fluctuations are no longer fully characterized by the power spectrum only because non-Gaussian features develop. A perturbation becomes non-linear when $\delta \sim 1$; this happens first to the small scale perturbations (large $k$) and then to the large scale perturbations (small $k$). It can be shown that $\sigma^2 \sim \delta^2$, so when $\delta \sim 1$ then $\sigma^2 \sim 1$. At each cosmological epoch it is possible to define a characteristic scale length, the *non-linear scale*, at which the variance of the matter density contrast becomes 1. This scale length separates the small-scale regime where the dark matter power spectrum is dominated by the contributions of presumably virialized halos (non-linear evolution is much *faster* than linear, the collapse of a structure can be considered as almost instantaneous after reaching the non-linear scale) from the large-scale regime where the dark-matter density can be considered as a linear superposition of linearly evolved perturbation modes.

## 1.3 On clusters of galaxies

Galaxies are not randomly distributed in the sky. Their positions are correlated, and there are areas in the sky where the galaxy density is noticeably higher or lower than average. There are groups consisting of a few galaxies, and there are clusters of galaxies in which some hundred up to a thousand galaxies appear very close together. Clusters of galaxies are the largest gravitationally bound objects today in the Universe and represent the high-mass end of collapsed structures. They are of particular interest for cosmology because their number density, individual properties and spatial distribution constrain the power spectrum of the density fluctuations. Their formation history is sensitive to the cosmological parameters of the Universe: if the matter density in the universe is high, clusters tend to form later in cosmic history than if the matter density is low. Compactness and morphology of clusters then reflect the cosmic matter density, and this fact has various observable implications.

Clusters are also interesting structures on their own. Zwicky noted in 1933 that the galaxies in the Coma cluster and other rich clusters move so fast that the clusters required about 10 to 100 times more mass to keep the galaxies bound than could be accounted for by the luminous galaxies themselves. This was the earliest indication that there is invisible mass, or dark matter, in at least some objects in the Universe.

When X–ray telescopes became available after 1966, it was discovered that clusters emit $\approx 10^{43 \div 45} \mathrm{erg\, s^{-1}}$. The source of this powerful X-ray emission (thermal bremsstrahlung, free-free radiation) is a hot, dilute plasma with temperatures in the range $10^{7 \div 8}$K and

densities of $10^{-3}$cm$^{-3}$. Based on the assumption that this intra-cluster gas is in hydrostatic equilibrium with a spherically symmetric gravitational potential of the total cluster matter, the X–ray temperature and flux can be used to estimate the cluster mass. Typical results approximately agree with the mass estimates from the kinematics of cluster galaxies employing the virial theorem. Today we know that about 80% of the cluster mass is dark matter, and about 15% is the intra-cluster medium (the hot gas). The X–ray emission thus independently confirms the existence of dark matter in galaxy clusters. Typical cluster masses are about $10^{14\div15}$M$_\odot$, typical sizes are of order several Mpc.

Later, luminous arc-like features were discovered in galaxy clusters. Their light is typically bluer than that from the cluster galaxies, and their length is comparable to the size of the central cluster region. These arcs are images of galaxies in the background of the clusters which are strongly distorted by the gravitational tidal field close to the cluster centers. Galaxy clusters thus can show *strong lensing* signatures near their cores, but at larger distances from the center they can only weakly deform images of background galaxies. The high number density of weakly distorted background images allows one to measure the coherent distortion caused by the tidal gravitational field of the cluster out to fairly large radii. This means that clusters of galaxies can also be used for *weak lensing studies*. One of the main applications of weak gravitational lensing is to reconstruct the (projected) mass distribution of galaxy clusters from their measurable tidal fields. Another application is to look for coherent distortions of galaxy images to detect the presence of a galaxy cluster. This is the idea behind the approach used in this thesis to detect cluster halos.

## 1.4   On correlation functions, power spectra and clustering studies

Let us consider a random field $g(\boldsymbol{x})$ whose expectation value is zero everywhere.

If we have a field whose average is not zero than we can just consider the difference between the field and its expectation value. Spatial positions $\boldsymbol{x}$ have in general $n$ dimensions, and the field can be either real or complex. A random field $g(\boldsymbol{x})$ is called homogeneous if it cannot statistically be distinguished from the field $g(\boldsymbol{x} + \boldsymbol{r})$, where $\boldsymbol{r}$ is an arbitrary translation vector. Similarly, a random field $g(\boldsymbol{x})$ is called isotropic if it has the same statistical properties as the random field $g(\mathscr{R}\boldsymbol{x})$, where $\mathscr{R}$ is an arbitrary rotation matrix in $n$ dimensions.

The *correlation function* of a random field is a measurement of the level of order of a system; it describes how microscopic variables are connected and measures how the field co-varies in average over space. It is defined as

$$\xi(r) = C_{gg}(|\boldsymbol{x} - \boldsymbol{x}'|) = \langle g(\boldsymbol{x})\, g^*(\boldsymbol{x}')\rangle, \tag{1.36}$$

and for a homogeneous and isotropic random field it depends only on the absolute value of the difference vector between the two points $\boldsymbol{x}$ and $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{r}$. The correlation function is averaged twice. First, we fix $\boldsymbol{x}$, we compute $g(\boldsymbol{x}) \, g^*(\boldsymbol{x}')$ for every point $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{r}$ that is distant $r = |\boldsymbol{r}|$ from $\boldsymbol{x}$ and average all these values. Then, we repeat for every $\boldsymbol{x}$ and average these values.

When the function $g$ is the density contrast field $\delta(\boldsymbol{x})$ defined in (1.26) the correlation function is the observable used in *clustering studies*. The correlation function is real, even when the field $g$ is complex. This can be seen by taking the complex conjugate of equation (1.36), which is equivalent to interchanging $\boldsymbol{x}$ and $\boldsymbol{x}'$, leaving the right-hand-side unaffected. We define the Fourier-transform pair of $g$ as

$$\tilde{g}(\boldsymbol{k}) = \int_{\mathbb{R}^n} g(\boldsymbol{x}) \, e^{-i\,\boldsymbol{k}\cdot\boldsymbol{x}} \, \mathrm{d}^n\boldsymbol{x}. \tag{1.37}$$

$$g(\boldsymbol{x}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \delta(\boldsymbol{x}) \, e^{i\,\boldsymbol{k}\cdot\boldsymbol{x}} \, \mathrm{d}^n\boldsymbol{k}. \tag{1.38}$$

The correlation function can be computed in the Fourier space as:

$$\langle \tilde{g}(\boldsymbol{k}) \, \tilde{g}^*(\boldsymbol{k}') \rangle = \int_{\mathbb{R}^n} e^{-i\,\boldsymbol{k}\cdot\boldsymbol{x}} \int_{\mathbb{R}^n} \langle g(\boldsymbol{x}) g^*(\boldsymbol{x}') \rangle \, e^{i\,\boldsymbol{k}'\cdot\boldsymbol{x}'} \, \mathrm{d}^n\boldsymbol{x} \, \mathrm{d}^n\boldsymbol{x}'. \tag{1.39}$$

Using equation (1.36) and by using $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{r}$, this becomes

$$
\begin{aligned}
\langle \tilde{g}(\boldsymbol{k}) \, \tilde{g}^*(\boldsymbol{k}') \rangle &= \int_{\mathbb{R}^n} e^{-i\,\boldsymbol{k}\cdot\boldsymbol{x}} \int_{\mathbb{R}^n} C_{gg}(|\boldsymbol{r}|) \, e^{i\,\boldsymbol{k}'\cdot(\boldsymbol{x}+\boldsymbol{r})} \, \mathrm{d}^n\boldsymbol{x} \, \mathrm{d}^n\boldsymbol{x}' = \\
&= (2\pi)^n \, \delta_D^{(n)}(\boldsymbol{k} - \boldsymbol{k}') \int_{\mathbb{R}^n} C_{gg}(|\boldsymbol{r}|) \, e^{-i\,\boldsymbol{k}\cdot\boldsymbol{r}} \, \mathrm{d}^n\boldsymbol{r} = \\
&= (2\pi)^n \, \delta_D^{(n)}(\boldsymbol{k} - \boldsymbol{k}') \, \mathcal{P}_g(k).
\end{aligned}
\tag{1.40}
$$

In the final step, we defined the power spectrum of the homogeneous and isotropic random field $g$ as

$$\mathcal{P}_g(k) = \int_{\mathbb{R}^n} C_{gg}(|\boldsymbol{r}|) \, e^{-i\,\boldsymbol{k}\cdot\boldsymbol{r}} \, \mathrm{d}^n\boldsymbol{r}, \tag{1.41}$$

which is the Fourier transform of the two-point correlation function. Isotropy of the random field implies that $\mathcal{P}_g$ can only depend on the modulus of $\boldsymbol{k}$.

Gaussian random fields are characterized by the property that the probability distribution of any linear combination of the random field $g(\boldsymbol{x})$ is Gaussian. The joint probability distribution of a number $M$ of linear combinations of the random variable $g(\boldsymbol{x_i})$ is a multivariate Gaussian. This is equivalent to requiring that the Fourier components $\tilde{g}(\boldsymbol{k})$ are mutually statistically independent, and that the probability densities for the $\tilde{g}(\boldsymbol{k})$ are Gaussian with dispersion $\mathcal{P}_g(k)$. Thus, a Gaussian random field is fully characterized by its power spectrum.

### 1.4.1   On the power spectrum used in Cosmology

In Cosmology and clustering studies, the statistical approach to perturbation theory is based on the *ergodic hypothesis*: the mean value of a given quantity when averaged on a large number of volumes (each one representing a realization of the Universe) is equal to the average of the mean values of the quantity averaged on many sub-volumes of a single realization of the Universe. For Gaussian probability distributions the ergodic hypothesis is always verified, and this is why the Gaussian distribution can easily be used in clustering studies. The sub-volumes over which mean values are computed must be large, independent, must be a good representation of the Universe (homogeneous and isotropic) and must be a fair sample. The probability distribution of $\delta$ is then normal with zero mean; it is described only by its variance $\sigma^2$:

$$P(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\delta^2}{2\sigma^2}\right). \tag{1.42}$$

This is an approximation, since for $\delta > -1$ the distribution must be truncated. The correlation function of the density contrast field $\delta(\boldsymbol{x})$ defined in equation (1.26) can be expressed as:

$$\xi(r) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \mathcal{P}(k)\, e^{i\,\boldsymbol{k}\cdot\boldsymbol{r}}\, \mathrm{d}^3\boldsymbol{k}. \tag{1.43}$$

Its Power spectrum is then:

$$\langle \tilde{\delta}(\boldsymbol{k})\, \tilde{\delta}^*(\boldsymbol{k'}) \rangle = (2\pi)^3\, \delta_D^{(3)}(\boldsymbol{k} - \boldsymbol{k'})\, \mathcal{P}(k), \tag{1.44}$$

it has the dimensions of a volume. The power spectrum is proportional to the average square amplitude of a perturbation with wave vector $k$. To be more accurate $\mathcal{P}(k)$ is a power density and $\mathcal{P}\, \mathrm{d}^3\boldsymbol{k}$ is the actual "power" i.e. the amplitude of the mode $k$ of the correlation function. Dirac's delta is not zero only when $\boldsymbol{k} = \boldsymbol{k'}$, and in that case

$$\mathcal{P}(k) \propto \langle \tilde{\delta}(\boldsymbol{k})\, \tilde{\delta}^*(\boldsymbol{k}) \rangle = \langle |\tilde{\delta}(\boldsymbol{k})|^2 \rangle. \tag{1.45}$$

The proportionality constant is a normalization factor known as *volume of the Universe* $V_\infty$. It is given expanding Dirac's delta, equation (1.31), as an integral when $\boldsymbol{k} = \boldsymbol{k'}$:

$$\delta_D^{(3)}(\boldsymbol{k} - \boldsymbol{k'}) = \delta_D^{(3)}(\boldsymbol{0}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \mathrm{d}^3\boldsymbol{x} \equiv \frac{V_\infty}{(2\pi)^3}. \tag{1.46}$$

The power spectrum can be written as $\mathcal{P}(k) = \dfrac{\langle |\tilde{\delta}(\boldsymbol{k})|^2 \rangle}{V_\infty}$. Let us also note that an average in the Fourier space means averaging every wave vector $\boldsymbol{k}$ whose wave number is $|\boldsymbol{k}| = k$.

The power spectrum is an extremely important quantity because it can give the variance of the probability distribution of $\delta$, determining the only parameter of that

Gaussian distribution. In fact, thanks to the ergodic hypothesis, the variance of $\delta$ can be expressed as an average of the values computed in a large number of universe sub-volumes:

$$\sigma^2 = \langle \delta^2(\boldsymbol{x}) \rangle = \frac{1}{V_\infty} \int_{\mathbb{R}^3} \langle \delta^2(\boldsymbol{x}) \rangle \, \mathrm{d}^3\boldsymbol{x} = \frac{1}{V_\infty} \int_{\mathbb{R}^3} \langle \delta(\boldsymbol{x}) \, \delta^*(\boldsymbol{x}) \rangle \, \mathrm{d}^3\boldsymbol{x} =$$

$$= \frac{1}{V_\infty (2\pi)^3} \int_{\mathbb{R}^3} \langle \tilde{\delta}(\boldsymbol{k}) \, \tilde{\delta}^*(\boldsymbol{k}) \rangle \, \mathrm{d}^3\boldsymbol{k} = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \frac{\langle |\tilde{\delta}(\boldsymbol{k})|^2 \rangle}{V_\infty} \, \mathrm{d}^3\boldsymbol{k} =$$

$$= \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \mathcal{P}(k) \, \mathrm{d}^3\boldsymbol{k} = \frac{1}{2\pi^2} \int_0^\infty \mathcal{P}(k) k^2 \, \mathrm{d}k.$$

Parseval's theorem was used in this demonstration. The variance is then the second momentum of the power spectrum:

$$\sigma^2 = \frac{1}{2\pi^2} \int_0^\infty \mathcal{P}(k) k^2 \, \mathrm{d}k. \tag{1.47}$$

As it was discussed in section 1.2, the initial power spectrum given by the inflationary expansion phase of the Universe is a power law, that has no privileged scale (scale-free):

$$\mathcal{P}(k) = A \, k^n \tag{1.48}$$

The amplitude $A$ is not predicted by models and must be constrained by observations. The inflation theory suggests $n = 1$, that is a scale-invariant spectrum known as *Zel'dovich Spectrum*. The evolution of the perturbations modifies the power spectrum in scale-dependent way, so that it can be usually described as a local power law with local index $n_{\mathrm{eff}}(k)$, defined as

$$n_{\mathrm{eff}}(k) = \frac{\mathrm{d} \ln \mathcal{P}(k)}{\mathrm{d}k}. \tag{1.49}$$

According to equation (1.34) the power spectrum evolves with time as

$$\mathcal{P}(k, t) = \mathcal{P}_i(k) \, \delta_+^2(t),$$

where the function $\delta_+(t)$ describes the time evolution and is equivalent to $a_{ent}^q$ used in equation (1.34).

Since the variance found in equation (1.47) is an integral of the power spectrum it also has a time evolution, that is given by:

$$\sigma^2 \propto \delta_+^2(t) \, k^{n+3} \propto \delta_+^2(t) \, \lambda^{-(n+3)} \propto \delta_+^2(t) \, M^{-\frac{n+3}{3}}. \tag{1.50}$$

This can be obtained by using the power law power spectrum of equation (1.48) in equation (1.47) and solving the integral. The variable $\lambda$ is the scale length of the perturbation that has wave vector $k = \frac{2\pi}{\lambda}$, and $M$ is the scale mass of the perturbation, given by

$M \propto \lambda^3 \propto k^{-3}$. As it was discussed above at page 14, when $\sigma^2 \sim 1$ the evolution of the perturbations changes from a linear regime into a non-linear regime. This happens when

$$M \sim \delta_+^{\frac{6}{n+3}}(t). \tag{1.51}$$

This is the *non-linear scale* expressed as a mass scale.

The Zel'dovich spectrum can be obtained by considering that the inflation causes perturbations in the spacetime metric, that become perturbations in the potential, that have no privileged scale and are constant (scale-invariant). For a fixed $R$ the fluctuations of the potential are:

$$\delta\Phi \propto \frac{G\delta M}{R} \propto \frac{G\delta\rho R^3}{R} \propto \delta\rho R^2 \propto \sigma R^2$$
$$\delta\Phi \propto \sigma M^{\frac{2}{3}} \propto M^{-\frac{n+3}{6}+\frac{2}{3}} \propto M^{\frac{1-n}{6}}$$

To have a constant amplitude it must be $n = 1$.

# Chapter 2

# An introduction to gravitational lensing

This chapter briefly reviews the main aspects of the theory of *Gravitational lensing*, with a particular emphasis on the weak-lensing regime. In this introduction to the chapter we are going to discuss what gravitational lensing is as a branch of Astrophysics. In section 2.1 we are going to discuss several quantities used in lensing studies and relevant for this thesis, such as the lens equation, convergence, shear, magnification, the E/B mode decomposition. In section 2.2 we focus on the description of a statistics known as the *Aperture mass*. The main references for this chapter are Bartelmann et al. (2001) and Umetsu (2020).

It is a prediction of Einstein's theory of General Relativity that light rays are deflected when they propagate through an inhomogeneous gravitational field. Several researchers before Einstein had speculated about such an effect, and assuming light behaves like a stream of particles, its deflection can be calculated even within Newton's theory of gravitation; General Relativity predicts deflections twice larger than they are in classical gravitation. A measure of the deflection of a Sun's light ray by $1.75''$, as predicted by General Relativity, was one of the most important step towards accepting General Relativity as the correct theory of gravity. The interest in gravitational lensing increased after the discovery of the first multiple-image system in 1979: distant and massive bodies can bend light rays from a source so strongly that multiple light rays can reach the observer (there may exist more than one null geodesic connecting the world-line of a source with the observation event). Another relevant lensing effect is magnification: gravitational lensing does not change the surface brightness of a source, but it bends the effective solid angle under which its light can be seen and increase its flux. This allows to use massive cosmic bodies (mainly clusters of galaxies) as "cosmic telescopes" to see very distant and faint sources. The images of resolved sources can also be deformed by lensing effects; this deformation can either be strong, as for gravitational arcs and Einstein

rings, or weak, like a change in the source's ellipticity. Although weak distortions in individual images can hardly be recognised, the net distortion averaged over an ensemble of images can still be detected. These different effects suggest the division of gravitational lensing into two regimes: *strong lensing* and *weak lensing.* Although there are not exact definitions, strong lensing deals with lensing effects produced by a mass distribution (the lens) over one or more background sources, so that their light is strongly distorted. Gravitational arcs and multiple images are common in strong lensing. Weak lensing has no such evident effects and in general cannot be attributed to a single lens (although there are exceptions); measurements of its effects are statistical in nature and show up only across ensembles of sources. Magnification and distortion effects due to weak lensing can be used to probe the statistical properties of the matter distribution between us and an ensemble of distant sources, provided some assumptions on the source properties can be made.

Another relevant property of gravitational lensing is that the deflection angle of a light ray is determined only by the gravitational field of the matter distribution along its path, regardless of the nature or physical state of the matter: light deflection probes the total matter density, without distinguishing between ordinary (baryonic) matter or dark matter. In contrast to other dynamical methods for probing gravitational fields, no assumption needs to be made on the dynamical state of the matter. For example, the interpretation of radial velocity measurements in terms of the gravitating mass requires the applicability of the virial theorem (i.e., the physical system is assumed to be in virial equilibrium), or knowledge of the orbits (such as the circular orbits in disk galaxies). However, lensing measures only the mass distribution projected along the line-of-sight, and is therefore insensitive to the extent of the mass distribution along the light rays, as long as this extent is small compared to the distances from the observer and the source to the deflecting mass. The possibility to directly investigate the dark-matter distribution led to substantial results over the years. Constraints on the size of the dark-matter haloes of spiral galaxies were derived, the presence of dark-matter haloes in elliptical galaxies was demonstrated, and the projected total mass distribution in many clusters of galaxies was mapped. These results directly impact on our understanding of structure formation, supporting hierarchical structure formation in cold dark matter (CDM) models. Constraints on the nature of dark matter were also obtained. Compact dark-matter objects, such as black holes or brown dwarfs, cannot be very abundant in the Universe, because otherwise they would lead to observable lensing effects. Galactic microlensing experiments constrained the density and typical mass scale of massive compact halo objects in our Galaxy. See Bartelmann et al. (2001) and Umetsu (2020) for more references.

## 2.1 On the basics of gravitational lensing

The presence of a point-like mass $M$ in position $\boldsymbol{\xi}'$ can deflect a ray of light (Bartelmann et al. 2001) by an angle

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4GM}{c^2} \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2}. \tag{2.1}$$

The vector $\boldsymbol{\xi}$ is the bidimensional position on a plane perpendicular to the line of sight. If we have a mass distribution whose surface density on the sky plane is $\Sigma(\boldsymbol{\xi}')$ then the deflection angle that a ray of light passing in position $\boldsymbol{\xi}$ in the plane is given by

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int_{\mathbb{R}^2} \Sigma(\boldsymbol{\xi}') \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2} \, \mathrm{d}^2\boldsymbol{\xi}'. \tag{2.2}$$

This expression is valid as long as the deviation of the actual light ray from a straight (undeflected) line within the mass distribution is small compared to the scale on which the mass distribution changes significantly. This condition is satisfied in virtually all astrophysically relevant situations (i.e. lensing by galaxies and clusters of galaxies), unless the deflecting mass extends all the way from the source to the observer. A notable exception, in particular for this thesis, is the deflection by the large scale structure.

A very important tool in lensing is the *lens equation*. Let us suppose we have a light source in position $\boldsymbol{\eta}$ in the source plane, that we would see at angular position $\boldsymbol{\beta}$ if there was no lens. Let us also suppose that we have a lens in the origin of the lens plane (it is possible to generalise). It deviates the light rays coming from the source, so that the observer gets to see the light ray passing through the lens plane in position $\boldsymbol{\xi}$ at angular position $\boldsymbol{\theta}$. Its position in the source plane is $\boldsymbol{\eta}'$.
This is shown in figure 2.1. We can define:

$$\hat{\boldsymbol{\eta}} \equiv \boldsymbol{\eta}' - \boldsymbol{\eta}. \tag{2.3}$$

We can define the angular diameter distances of the lens $D_L$, of the source $D_S$, the lens-source angular distance $D_{LS}$ as

$$\boldsymbol{\xi} \equiv D_L \, \boldsymbol{\theta}, \tag{2.4}$$
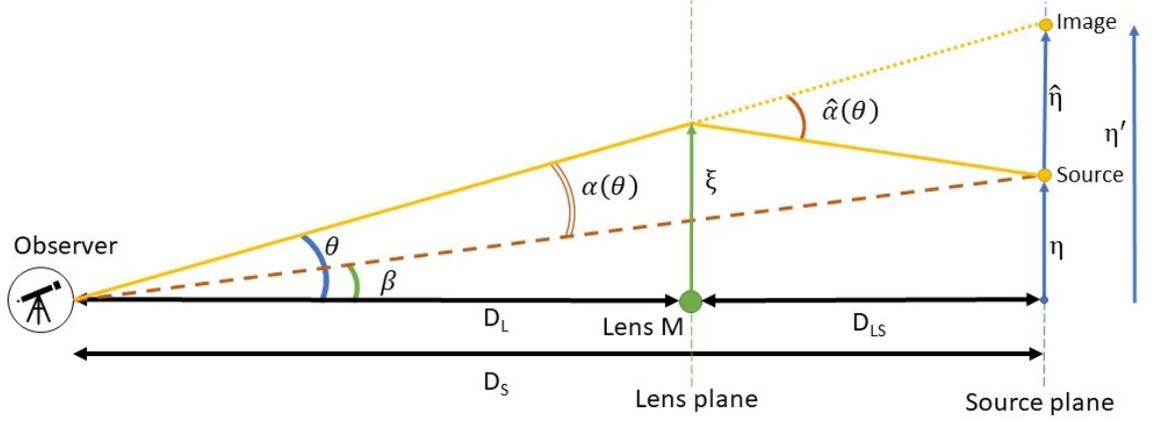$$\boldsymbol{\eta} \equiv D_S \, \boldsymbol{\beta}, \tag{2.5}$$
$$\hat{\boldsymbol{\eta}} \equiv D_{LS} \, \hat{\boldsymbol{\alpha}}, \tag{2.6}$$

and note that

$$\boldsymbol{\eta}' = D_S \, \boldsymbol{\theta}. \tag{2.7}$$

In general $D_S \neq D_L + D_{LS}$, and the equality holds only in the local Universe, where we can ignore the expansion of the Universe and the geometry is Euclidean. Putting the definitions (2.5),(2.6), (2.7) in (2.3) we have:

$$D_{LS} \, \hat{\boldsymbol{\alpha}} = D_S \, \boldsymbol{\theta} - D_S \, \boldsymbol{\beta},$$

**Figure 2.1:** Graphical description of the observer-lens-source system assumed to derive the lens equation. The source and lens planes are projected.

from which we can obtain the *lens equation*

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{2.8}$$

where we have defined the *Reduced deflection angle*

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) \equiv \frac{D_{LS}}{D_S} \, \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}). \tag{2.9}$$

A useful lensing quantity is the *lensing potential* of the gravitational lens. If we consider that the three-dimensional position of the lens $\boldsymbol{\xi} + \boldsymbol{x_{los}}$ is the sum of a vector $\boldsymbol{\xi} = D_L \boldsymbol{\theta}$ on the lens plane plus a vector $\boldsymbol{x_{los}}$ on the *line of sight*, we can define the lensing potential as a scaled integral of the gravitational potential $\phi$ of the lens on the line of sight:

$$\psi(\boldsymbol{\theta}) \equiv \frac{D_{LS}}{D_L \, D_S} \frac{2}{c^2} \int_{\mathbb{R}} \phi \left( D_L \, \boldsymbol{\theta}, x_{los} \right) \, \mathrm{d}x_{los}. \tag{2.10}$$

The lensing potential is adimensional. It can be shown that:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \, \psi(\boldsymbol{\theta}). \tag{2.11}$$

If we define the *critical surface density* of a source-lens system

$$\Sigma_{crit} \equiv \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}} \tag{2.12}$$

we can define the *convergence $\kappa$* of the system as an adimensional surface density:

$$\kappa(\boldsymbol{\theta}) \equiv \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_{crit}}. \tag{2.13}$$

24

It can be shown that:

$$\nabla_{\boldsymbol{\theta}}^2 \psi(\boldsymbol{\theta}) = 2\kappa(\boldsymbol{\theta}). \tag{2.14}$$

This is the Poisson equation projected on the lens plane.

Another way to express the reduced deflection angle is the following:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \kappa(\boldsymbol{\theta}') \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} \, \mathrm{d}^2\boldsymbol{\theta}'. \tag{2.15}$$

This expression can be obtained by rewriting equation (2.9) with equations (2.2), (2.4) and (2.12). If we use equation (2.15) in equation (2.11) and integrate the equation we can express the lensing potential as a convolution:

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \kappa(\boldsymbol{\theta}') \ln|\boldsymbol{\theta} - \boldsymbol{\theta}'| \, \mathrm{d}^2\boldsymbol{\theta}'. \tag{2.16}$$

### 2.1.1 On the first-order lens mapping

Gravitational lensing causes a displacement of the apparent position of the source as described by the lens equation (2.8). If a source is much smaller than the angular scale on which the lens properties change, the lens mapping can be locally linearised. Let us suppose we have a source in position $\boldsymbol{\beta}$, an image with position $\boldsymbol{\theta}$ and deflection $\boldsymbol{\alpha}$, and let us suppose we move the source by a little, so that we now have $\boldsymbol{\theta}'$, $\boldsymbol{\beta}'$ and $\boldsymbol{\alpha}'$. Then we can consider a first order approximation of the lens mapping:

$$\mathrm{d}\boldsymbol{\beta} = \boldsymbol{\beta}' - \boldsymbol{\beta} = \boldsymbol{\theta}' - \boldsymbol{\alpha}' - \boldsymbol{\theta} + \boldsymbol{\alpha} \simeq \mathrm{d}\boldsymbol{\theta} + \boldsymbol{\alpha} + \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\theta}} \mathrm{d}\boldsymbol{\theta} - \boldsymbol{\alpha}' = \left[ \mathbb{1} - \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\theta}} \right] \mathrm{d}\boldsymbol{\theta}, \tag{2.17}$$

where we have done a first-order Taylor expansion on $\boldsymbol{\alpha}'$. A similar expansion on $\boldsymbol{\beta}'$ would bring us to write

$$\mathrm{d}\boldsymbol{\beta} = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} \, \mathrm{d}\boldsymbol{\theta}. \tag{2.18}$$

This allows us to define the *Lensing Jacobian Matrix*

$$\mathcal{A} \equiv \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \left[ \mathbb{1} - \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\theta}} \right]. \tag{2.19}$$

By using equation (2.11) its elements can be expressed as second derivatives of the lensing potential:

$$\mathcal{A}_{ij}(\boldsymbol{\theta}) = \delta_{ij} - \frac{\partial^2 \psi(\boldsymbol{\theta})}{\partial \theta_i \, \partial \theta_j} = \delta_{ij} - \psi_{ij}(\boldsymbol{\theta}), \tag{2.20}$$

where $\psi_{ij} \equiv \dfrac{\partial^2 \psi(\boldsymbol{\theta})}{\partial \theta_i \, \partial \theta_j}$.

The lensing Jacobian is a symmetric tensor, so it can be decomposed as the sum of an isotropic part and an anisotropic one. The isotropic part, using the Poisson equation (2.14) can be expressed as a function of the convergence only:

$$\mathcal{A}_{ij}^{iso}(\boldsymbol{\theta}) = \frac{1}{2} \operatorname{Tr} \mathcal{A}(\boldsymbol{\theta}) \, \delta_{ij} = \frac{1}{2} \left[ 2 - (\psi_{11} + \psi_{22}) \right] \delta_{ij} = [1 - \kappa(\boldsymbol{\theta})] \delta_{ij}. \tag{2.21}$$

The effect of an isotropic deflection is then to scale vectors in the source plane by a factor $1 - \kappa(\boldsymbol{\theta})$. Let us also note that the convergence is

$$\kappa(\boldsymbol{\theta}) = \frac{1}{2} [\psi_{11}(\boldsymbol{\theta}) + \psi_{22}(\boldsymbol{\theta})]. \tag{2.22}$$

The anisotropic part of the lensing Jacobian is:

$$\mathcal{A}_{ij}^{aniso}(\boldsymbol{\theta}) = \mathcal{A}_{ij} - \frac{1}{2} \operatorname{Tr} \mathcal{A} \, \delta_{ij} = \delta_{ij} - \psi_{ij} - (1 - \kappa)\delta_{ij}. \tag{2.23}$$

The opposite of this matrix is known as *shear tensor* $\Gamma$:

$$\Gamma_{ij}(\boldsymbol{\theta}) = -\mathcal{A}_{ij}^{aniso}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{2}(\psi_{11} - \psi_{22}) & \psi_{12} \\ \psi_{12} & -\frac{1}{2}(\psi_{11} - \psi_{22}) \end{bmatrix} = \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{bmatrix}. \tag{2.24}$$

It is a symmetric tensor with null trace with two independent *shear components* $\gamma_1$ and $\gamma_2$:

$$\gamma_1(\boldsymbol{\theta}) = \frac{1}{2} \left[ \psi_{11}(\boldsymbol{\theta}) - \psi_{22}(\boldsymbol{\theta}) \right], \tag{2.25}$$

$$\gamma_2(\boldsymbol{\theta}) = \psi_{12}(\boldsymbol{\theta}). \tag{2.26}$$

We can define the *shear* as the complex number $\gamma = \gamma_1 + i\,\gamma_2$, its *shear modulus* $|\gamma| = \sqrt{\gamma_1^2 + \gamma_2^2}$, then $\det \Gamma = -|\gamma|^2$ and the eigenvalues of the shear tensor are $\pm|\gamma|$. This decomposition of the lensing Jacobian into a convergence-dependent, isotropic part and a shear-dependent, anisotropic part allows us to write:

$$\mathcal{A}_{ij} = \mathcal{A}_{ij}^{iso} + \mathcal{A}_{ij}^{aniso} = [1 - \kappa(\boldsymbol{\theta})]\delta_{ij} - \Gamma_{ij} = \begin{bmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{bmatrix}. \tag{2.27}$$

This tensor is real and symmetric, so it can be diagonalised. Its determinant is:

$$\det \mathcal{A}(\boldsymbol{\theta}) = [1 - \kappa(\boldsymbol{\theta}) - |\gamma|(\boldsymbol{\theta})] \, [1 - \kappa(\boldsymbol{\theta}) + |\gamma|(\boldsymbol{\theta})]. \tag{2.28}$$

Its eigenvalues are:

$$\Lambda_+(\boldsymbol{\theta}) = 1 - \kappa(\boldsymbol{\theta}) + |\gamma(\boldsymbol{\theta})|, \tag{2.29}$$

$$\Lambda_-(\boldsymbol{\theta}) = 1 - \kappa(\boldsymbol{\theta}) - |\gamma(\boldsymbol{\theta})|. \tag{2.30}$$

The first-order lens mapping formalism here described allows us to write the lens equation as:

$$\boldsymbol{\beta} = \mathcal{A}(\boldsymbol{\theta})\,\boldsymbol{\theta}. \tag{2.31}$$

The position of the source and the position of the image are connected by a linear map. The linear map describes the deflection of light rays when they pass from the source plane to the lens plane. The total deflection is the sum of the deflections on the eigenvectors of $\mathcal{A}$ (and $\Gamma$), with different amplitudes given by the eigenvalues of $\mathcal{A}$.

## 2.1.2  On magnification and lensing regimes

The solutions $\boldsymbol{\theta}$ of the lens equation (2.8) yield the angular positions of the images of a source at $\boldsymbol{\beta}$. The shapes of the images will differ from the shape of the source because light bundles are deflected differentially. The most visible consequence of this distortion is the occurrence of giant luminous arcs in galaxy clusters. In general, the shapes of the images must be determined by solving the lens equation for all points within an extended source. Liouville's theorem and the absence of emission and absorption of photons in gravitational light deflection imply that lensing conserves surface brightness (or specific intensity). Hence, if $I^{(s)}(\boldsymbol{\beta})$ is the surface brightness distribution in the source plane, the observed surface brightness distribution in the lens plane is

$$I(\boldsymbol{\theta}) = I^{(s)}(\boldsymbol{\beta}). \tag{2.32}$$

When it is possible to use the linear approximation of the lens equation (2.31) if we have a point $\boldsymbol{\theta_0}$ within an image, corresponding to the point $\boldsymbol{\beta_0} = \boldsymbol{\beta}(\boldsymbol{\theta_0})$ within the source, we find from equation (2.32) using equation (2.31)

$$I(\boldsymbol{\theta}) = I^{(s)}\left[\boldsymbol{\beta_0} + \mathcal{A}(\boldsymbol{\theta_0})\left(\boldsymbol{\theta} - \boldsymbol{\theta_0}\right)\right]. \tag{2.33}$$

According to this equation, the images of a circular source are ellipses. The ratios of the semi-axes of such an ellipse to the radius of the source are given by the inverse of the eigenvalues of $\mathcal{A}(\boldsymbol{\theta_0})$, and the ratio of the solid angles subtended by an image and the unlensed source is the inverse of the determinant of $\mathcal{A}$. The fluxes observed from the image and from the unlensed source are given as integrals over the brightness distributions $I(\boldsymbol{\theta})$ and $I^{(s)}(\boldsymbol{\beta})$, respectively, and their ratio is called the *magnification* $\mu(\boldsymbol{\theta_0})$. From equation (2.28) we obtain:

$$\mu(\boldsymbol{\theta}) = \frac{1}{\det \mathcal{A}(\boldsymbol{\theta})} = \frac{1}{[1 - \kappa(\boldsymbol{\theta}) - |\gamma|(\boldsymbol{\theta})]\,[1 - \kappa(\boldsymbol{\theta}) + |\gamma|(\boldsymbol{\theta})]}. \tag{2.34}$$

The images are thus distorted in shape and size. The shape distortion is due to the tidal gravitational field, described by the shear $\gamma$, whereas the magnification is caused by both

isotropic focusing caused by the local matter density $\kappa$ and anisotropic focusing caused by shear.

Points in the lens plane where the Jacobian $\mathcal{A}$ is singular, i.e. where $\det(\mathcal{A}) = 0$, form closed curves called *critical curves*. Those are the curves where the eigenvalues (2.29) and (2.30) of the Jacobian become 0. The inner critical curve is the one where the eigenvalue $\Lambda_+(\boldsymbol{\theta}) = 0$; it is called the *radial critical curve* because images forming along it are stretched in the direction perpendicular to the curve. The outer critical curve is the one where the eigenvalue $\Lambda_-(\boldsymbol{\theta}) = 0$; it is called the *tangential critical curve* because images forming along it are distorted tangentially to the curve. The image curves of the critical lines in the source plane are called *caustics*. Equation (2.34) predicts that sources on caustics are infinitely magnified; however, infinite magnification does not occur in reality, for two reasons. First, each astrophysical source is extended, and its magnification (given by the surface brightness-weighted point-source magnification across its solid angle) remains finite. Second, even point sources would be magnified by a finite value since for them, the geometrical-optics approximation fails near critical curves, and a wave-optics description leads to a finite magnification (the finite-size effect always dominates). Nevertheless, images near critical curves can be magnified and distorted substantially, as is demonstrated by the giant luminous arcs which are formed from source galaxies close to caustics. When a source moves across a caustic its number of images, i.e. solutions of the lens equation (2.8), changes by 2, and the two additional images appear or disappear at the corresponding critical curve in the lens plane. Hence, only sources inside a caustic have multiple images.

A lens system that has a region with $\kappa(\boldsymbol{\theta}) > 1$ can produce multiple images for certain source positions $\beta$, and such a system is referred to as being *supercritical*, because the surface density is greater than the critical density. Being supercritical is a sufficient but not a necessary condition for a general lens to produce multiple images, because the shear can also contribute to multiple imaging. Nevertheless, this provides us with a simple criterion to broadly distinguish the regimes of multiple and single imaging. Keeping this in mind, we refer to the region where $\kappa(\boldsymbol{\theta}) \gtrsim 1$ as the *strong-lensing regime* and the region where $\kappa(\boldsymbol{\theta}) \ll 1$ as the *weak-lensing regime*.

### 2.1.3   On the properties of the shear tensor

The shear tensor (2.24) is symmetric and trace-less. It can be expressed in terms of the Pauli matrices as

$$\Gamma(\boldsymbol{\theta}) = \gamma_1(\boldsymbol{\theta})\,\sigma_3 + \gamma_2(\boldsymbol{\theta})\,\sigma_1. \tag{2.35}$$

It can also be expressed in index notation as:

$$\Gamma_{ij}(\boldsymbol{\theta}) = \left(\partial_i\partial_j - \delta_{ij}\,\frac{1}{2}\,\nabla^2\right)\psi(\boldsymbol{\theta}), \tag{2.36}$$

where $\partial_i = \dfrac{\partial}{\partial \theta_i}$, $\nabla^2 = \partial_i^2 + \partial_j^2$, $\delta_{ij}$ is Kronecker's delta.

If we rotate our coordinate system by an angle $\varphi$ thanks to a rotation matrix $R(\varphi)$ it can be shown that the rotation has no effect on the isotropic part of the matrix, the shear matrix is modified according to $\Gamma' = R^T(\varphi)\,\Gamma\,R(\varphi)$ and the shear components change as follows:

$$\begin{bmatrix} \gamma_1' \\ \gamma_2' \end{bmatrix} = R^T(2\varphi) \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \tag{2.37}$$

or, in a more explicit way:

$$\begin{cases} \gamma_1' = +\gamma_1 \cos(2\varphi) + \gamma_2 \sin(2\varphi) \\ \gamma_2' = -\gamma_1 \sin(2\varphi) + \gamma_2 \cos(2\varphi) \end{cases}. \tag{2.38}$$

The shear components are mapped on themselves after a $\pi$ rotation, so the shear is not a vector field, but a spin-2 field[1]. The Jacobian matrix can be diagonalised with a rotation that makes $\gamma_1' = \gamma$ and $\gamma_2' = 0$, which can be obtained if $\varphi = \frac{1}{2}\arctan\left(\frac{\gamma_2}{\gamma_1}\right)$.

Let us suppose we have a circular source centered in $\boldsymbol{\beta} = \mathbf{0}$ with radius $\beta$. This means that $\beta^2 = \beta_1^2 + \beta_2^2$. Let us also suppose that our coordinate axes are the eigenvectors. Equation (2.31) allows us to express $\beta_1$ and $\beta_2$ as a function of $\theta_1$ and $\theta_2$:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1-\kappa-|\gamma| & 0 \\ 0 & 1-\kappa+|\gamma| \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} (1-\kappa-|\gamma|)\,\theta_1 & 0 \\ 0 & (1-\kappa+|\gamma|)\,\theta_2 \end{bmatrix}.$$

This means that the circularity relation becomes:

$$\beta^2 = \beta_1^2 + \beta_2^2 = (1-\kappa-|\gamma|)^2\,\theta_1^2 + (1-\kappa+|\gamma|)^2\,\theta_2^2,$$

from which we find out the image of the source is not circular, but an ellipse:

$$\frac{\theta_1^2}{\left(\frac{\beta}{1-\kappa-|\gamma|}\right)^2} + \frac{\theta_2^2}{\left(\frac{\beta}{1-\kappa+|\gamma|}\right)^2} = 1.$$

The circular source is mapped in an elliptical map whose axes are directed along the eigenvectors and their semiaxis lengths are different because of the shear. If $a$ and $b$ are the semiaxis lengths, the ellipticity $e$ of this ellipse is:

$$e = \frac{a+b}{a-b} = \frac{\frac{\beta}{1-\kappa-|\gamma|} + \frac{\beta}{1-\kappa+|\gamma|}}{\frac{\beta}{1-\kappa-|\gamma|} - \frac{\beta}{1-\kappa+|\gamma|}} = \frac{1-\kappa+|\gamma|-1+\kappa+|\gamma|}{1-\kappa+|\gamma|+1-\kappa-|\gamma|} = \frac{|\gamma|}{1-\kappa}. \tag{2.39}$$

This is equal to a quantity called *reduced shear modulus*:

$$|g| = \frac{|\gamma|}{1-\kappa}. \tag{2.40}$$

---

[1]A quantity is said to have spin $N$ if it has the same value after rotation by $\frac{2\pi}{N}$.

In general we can define the *Reduced shear tensor* as a scaled version of the shear tensor:

$$g(\boldsymbol{\theta}) = \frac{1}{1 - \kappa(\boldsymbol{\theta})} \, \Gamma(\boldsymbol{\theta}), \tag{2.41}$$

and by analogy its components $g_1$ and $g_2$ and the complex *reduced shear* $g = g_1 + i \, g_2$. The reduced shear is a fundamental quantity because if we measure the ellipticity of an intrinsic circular source then the observed ellipticity is a direct measure of the reduced shear and is an effect of the gravitational lensing only. Real galaxies have also intrinsic ellipticities, that makes everything more complex. Let us note that in the weak lensing limit $\kappa \ll 1$ we have $g \simeq \gamma$.

Another property of the complex shear is that it can be expressed as a convolution of the convergence just like the potential and the deflection angle by using (2.16):

$$\gamma = \gamma_1 + i \, \gamma_2 = \frac{1}{2}[\psi_{11} - \psi_{22}] + i \, \psi_{12} = \left[ \frac{1}{2} \frac{\partial^2}{\partial \theta_1^2} - \frac{1}{2} \frac{\partial^2}{\partial \theta_2^2} + i \, \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \right] \psi =$$

$$= \frac{1}{\pi} \int_{\mathbb{R}^2} \kappa(\boldsymbol{\theta}') \left[ \frac{1}{2} \frac{\partial^2}{\partial \theta_1^2} - \frac{1}{2} \frac{\partial^2}{\partial \theta_2^2} + i \, \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \right] \ln|\boldsymbol{\theta} - \boldsymbol{\theta}'| \, \mathrm{d}^2\boldsymbol{\theta}'.$$

Let us note that, with $i = 1, 2$:

$$\frac{\partial}{\partial \theta_i} \ln|\boldsymbol{\theta} - \boldsymbol{\theta}'| = \frac{1}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|} \frac{\partial}{\partial \theta_i} |\boldsymbol{\theta} - \boldsymbol{\theta}'| = \frac{\theta_i - \theta_i'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2},$$

$$\frac{\partial^2}{\partial \theta_i^2} \ln|\boldsymbol{\theta} - \boldsymbol{\theta}'| = \frac{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2 - 2(\theta_i - \theta_i')^2}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^4},$$

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln|\boldsymbol{\theta} - \boldsymbol{\theta}'| = -2\frac{(\theta_i - \theta_i')(\theta_j - \theta_j')}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^4}.$$

This brings us to:

$$\gamma(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \kappa(\boldsymbol{\theta}') \, \mathcal{D}(\boldsymbol{\theta} - \boldsymbol{\theta}') \, \mathrm{d}^2\boldsymbol{\theta}', \tag{2.42}$$

with

$$\mathcal{D}(\boldsymbol{\theta}) = \frac{\theta_2^2 - \theta_1^2 - 2 \, i \, \theta_1 \theta_2}{|\boldsymbol{\theta}|^4} = -\frac{1}{(\theta_1 - i \, \theta_2)^2}. \tag{2.43}$$

Its complex conjugate is

$$\mathcal{D}^*(\boldsymbol{\theta}) = \frac{\theta_2^2 - \theta_1^2 + 2 \, i \, \theta_1 \theta_2}{|\boldsymbol{\theta}|^4} = -\frac{1}{(\theta_1 + i \, \theta_2)^2}. \tag{2.44}$$

Another relation between shear and convergence can be found if we consider that they are both linear combinations of the second derivatives of the potential. In particular, if we introduce the complex *spin raising operator*

$$\partial = \partial_1 + i \, \partial_2 \tag{2.45}$$

and the complex *spin lowering operator*

$$\partial^* = \partial_1 - i\,\partial_2 \tag{2.46}$$

then $\nabla^2 = \partial_1^2 + \partial_2^2 = \partial\,\partial^*$. By using equations (2.22), (2.25) and (2.26) simple algebric calculations show that

$$\kappa = \frac{1}{2}\,\partial\,\partial^*\psi(\boldsymbol{\theta}) \tag{2.47}$$

$$\gamma = \frac{1}{2}\,\partial\,\partial\psi(\boldsymbol{\theta}) \tag{2.48}$$

$$\nabla^2\kappa(\boldsymbol{\theta}) = \partial^*\,\partial^*\gamma(\boldsymbol{\theta}) \tag{2.49}$$

### 2.1.4 On the E and B mode decomposition

Equation (2.35) showed that the shear can be decomposed thanks to two Pauli matrices, so it can be written as the sum of two linearly independent components. It is possible to introduce two scalar fields $\psi_E(\boldsymbol{\theta})$ and $\psi_B(\boldsymbol{\theta})$ so that the shear matrix can be decomposed into two independent modes E and B:

$$\Gamma(\boldsymbol{\theta}) = \Gamma^{(E)}(\boldsymbol{\theta}) + \Gamma^{(B)}(\boldsymbol{\theta}), \tag{2.50}$$

with

$$\Gamma_{ij}^{(E)}(\boldsymbol{\theta}) = \left(\partial_i\partial_j - \delta_{ij}\nabla^2\right)\psi_E(\boldsymbol{\theta}), \tag{2.51}$$

$$\Gamma_{ij}^{(B)}(\boldsymbol{\theta}) = \frac{1}{2}\left(\epsilon_{kj}\partial_i\partial_k + \epsilon_{ki}\partial_j\partial_k\right)\psi_B(\boldsymbol{\theta}), \tag{2.52}$$

where $\partial_i = \dfrac{\partial}{\partial\theta_i}$, $\nabla^2 = \partial_i^2 + \partial_j^2$, $\delta_{ij}$ is *Kronecker's delta*, $\epsilon_{ij}$ is the *Levi-Civita symbol* defined in two dimensions such as $\epsilon_{11} = \epsilon_{22} = 0$, $\epsilon_{12} = \epsilon_{21} = 1$. The decomposition is possible because, in general, any symmetric and trace-less bidimensional tensor field (such as the shear and the radiation field of electromagnetism) can be written as the sum of a curl-free term and a divergence-free term. In analogy to the radiation field of electromagnetism those are referred to as the "electric" (E) curl-free component and the "magnetic" (B) divergence-free component. The shear components (2.25) and (2.26) can be expressed as:

$$\gamma_1 = \frac{1}{2}(\psi_{E,11} - \psi_{E,22}) - \psi_{B,12}, \tag{2.53}$$

$$\gamma_2 = \psi_{E,12} + \frac{1}{2}(\psi_{B,11} - \psi_{B,22}). \tag{2.54}$$

If we compare equations (2.51) with (2.36), (2.53) with (2.25) and (2.54) with (2.26), we find that $\psi_E = \psi$ and $\psi_B = 0$: for a lensing-induced signal the E-mode signal is

related to the convergence $\kappa$ while the B-mode signal is identically null. The E/B mode decomposition was introduced (Crittenden et al. 2002) as a way to distinguish between the lensing and intrinsic ellipticities contributions to the observed ellipticity of an image. In fact equation (2.39) proved that the ellipticity of a circular source is the reduced shear, but this is not true in general if the source is intrinsically elliptical. If we want to use the measured ellipticity of an image as an estimate of its reduced shear we must keep in mind that the measured value is a combination of the intrinsic ellipticity of the source and the lensing effect. The lensing shear is curl-free, so its contribution to the ellipticity has only E-modes. In general lensing by a point mass will create a tangential, curl-free distortion pattern, so the most general distortion field produced by lensing will be a linear superposition of such patterns and will also be a curl-free field. However, the distortion field resulting from intrinsic spin alignments has E- and B-type modes of the same order of magnitude. This provides a powerful discriminant between lensing and intrinsic contributions to observed ellipticity correlations.

Umetsu (2020) shows that gravitational lensing can induce B modes, for example, when multiple deflections of light are involved (there are several lenses on different lens planes across the line of sight). However, these B modes can be generated at higher orders and the B-mode contributions coming from multiple deflections are suppressed by a large factor compared to the E-mode contributions. In real observations, intrinsic ellipticities of background galaxies also contribute to weak-lensing shear estimates. Assuming that intrinsic ellipticities have random orientations in projection space, such an isotropic ellipticity distribution will yield statistically identical contributions to the E and B modes. Therefore, the B-mode signal provides a useful null test for systematic effects in weak-lensing observations.

In chapter 3 equations (3.11), continuous version, and (3.18), discrete version, are going to introduce a statistics of galaxies ellipticities capable of extracting their E-mode. It will be used as an estimator of the weak lensing signal produced by dark matter haloes, but since real galaxies have intrinsic ellipticities it will evaluate also that contribution; the signal we want to observe and use to detect haloes will then be affected by a noise that comes from the intrinsic ellipticities (chapter 3 introduces other sources of noise too). A way to evaluate the "noisiness" of the map will be described by using a statistics of the B-mode only described in equation (3.19), that is not affected by the lensing signal but only by the galaxy ellipticities, considered as noise.

## 2.2 On the aperture mass

Equation (2.42) proved that the shear can be obtained as a convolution of the convergence. It can be inverted so that we can express the convergence as a convolution of the shear (Bartelmann et al. 2001; Kaiser et al. 1993; Schneider 1996).

Let us start by considering the Fourier transform of the convergence $\kappa$ and the shear convolution kernel $\mathcal{D}$ defined in (2.43):

$$\tilde{\kappa}(\boldsymbol{l}) = \int_{\mathbb{R}^2} e^{i\,\boldsymbol{\theta}\cdot\boldsymbol{l}}\,\kappa(\boldsymbol{\theta})\,\mathrm{d}^2\boldsymbol{\theta}, \tag{2.55}$$

$$\tilde{\mathcal{D}}(\boldsymbol{l}) = \pi\,\frac{l_1^2 - l_2^2 + 2i\,l_1 l_2}{|\boldsymbol{l}|^2}, \tag{2.56}$$

$$\tilde{\mathcal{D}}^*(\boldsymbol{l}) = \pi\,\frac{l_1^2 - l_2^2 - 2i\,l_1 l_2}{|\boldsymbol{l}|^2}, \tag{2.57}$$

let us note that $\tilde{\mathcal{D}}\,\tilde{\mathcal{D}}^* = \pi^2$ where the asterisk denotes the complex conjugate.

It is a known result that the Fourier transform of a convolution of two functions is the product of the Fourier transforms of the two functions, so $\tilde{\gamma}(\boldsymbol{l}) = \pi^{-1}\,\tilde{\kappa}(\boldsymbol{l})\,\tilde{\mathcal{D}}(\boldsymbol{l})$, from which we get $\tilde{\kappa}(\boldsymbol{l}) = \dfrac{1}{\pi}\,\tilde{\gamma}(\boldsymbol{l})\,\tilde{\mathcal{D}}^*(\boldsymbol{l})$. In the real space this becomes:

$$\kappa(\boldsymbol{\theta}) - \kappa_0 = \frac{1}{\pi}\int_{\mathbb{R}^2}\gamma(\boldsymbol{\theta}')\,\mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}')\,\mathrm{d}^2\boldsymbol{\theta}' = \frac{1}{\pi}\int_{\mathbb{R}^2}\Re\left[\gamma(\boldsymbol{\theta}')\,\mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}')\right]\,\mathrm{d}^2\boldsymbol{\theta}'. \tag{2.58}$$

The constant $\kappa_0$ shows that the shear does not change if we add a sheet of constant convergence. The symbol $\Re$ refers to the real part of the complex number, and the equality comes from the constraint that the convergence is a real function. In fact it can be shown that the imaginary part is $\Im\left[\gamma(\boldsymbol{\theta}')\,\mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}')\right] = 0$. One could think of using this relation to estimate the convergence from the shear by using a dataset of measured complex ellipticities $\epsilon_i = \epsilon(\boldsymbol{\theta_i})$ instead of the shear (equivalent to the reduced shear in the weak-lensing limit), and the estimator:

$$\kappa(\boldsymbol{\theta}) = \frac{1}{N\,\pi}\sum_{i=1}^{N}\Re\left[\epsilon_i\,\mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta_i})\right]. \tag{2.59}$$

This is one of the most basic estimator to reconstruct mass surface density from shear measurements and analyze it to detect mass concentrations, but the relations between galaxy ellipticities and the resulting mass maps are generally complicated.

A more common estimator is the *Aperture Mass*

$$M_{ap}(\boldsymbol{\theta}) = \int_{\mathbb{R}^2}\kappa(\boldsymbol{\theta}')\,U(\boldsymbol{\theta} - \boldsymbol{\theta}')\,\mathrm{d}^2\boldsymbol{\theta}'. \tag{2.60}$$

The aperture mass is the convolution of a convergence map with a weight function $U(\boldsymbol{\theta} - \boldsymbol{\theta}')$; this can be chosen as an "aperture function" that tends to zero when the distance $|\boldsymbol{\theta} - \boldsymbol{\theta}'|$ increases. In this way the aperture mass becomes an average estimate of the convergence weighted within the aperture with the weight function $U$. Let us assume

that the aperture is symmetric $U(\boldsymbol{\theta} - \boldsymbol{\theta}') = U(|\boldsymbol{\theta} - \boldsymbol{\theta}'|)$. We can insert equation (2.58) in equation (2.60) so it becomes:

$$M_{ap}(\boldsymbol{\theta}) = \int_{\mathbb{R}^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta}'|) \left[ \frac{1}{\pi} \int_{\mathbb{R}^2} \Re\left[ \gamma(\boldsymbol{\theta}'') \mathcal{D}^*(\boldsymbol{\theta}' - \boldsymbol{\theta}'') \right] \mathrm{d}^2\boldsymbol{\theta}'' + \kappa_0 \right] \mathrm{d}^2\boldsymbol{\theta}'$$

$$= \frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta}'') \int_{\mathbb{R}^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta}'|) \mathcal{D}^*(\boldsymbol{\theta}' - \boldsymbol{\theta}'') \mathrm{d}^2\boldsymbol{\theta}' \mathrm{d}^2\boldsymbol{\theta}'' + \kappa_0 \int_{\mathbb{R}^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta}'|) \mathrm{d}^2\boldsymbol{\theta}'.$$

The second integral can be written into polar coordinates as

$$\kappa_0 \int_{\mathbb{R}^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta}'|) \mathrm{d}^2\boldsymbol{\theta}' = \kappa_0 \, 2\pi \int_0^\infty U(x) x \, \mathrm{d}x.$$

Let us suppose that the weight function $U$ is *compensated*, that is

$$\int_0^\infty U(x) x \, \mathrm{d}x = 0. \tag{2.61}$$

In this way the aperture mass will not be dependent on $\kappa_0$; it can be expressed as:

$$M_{ap}(\boldsymbol{\theta}) = \frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta}'') \int_{\mathbb{R}^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta}'|) \mathcal{D}^*(\boldsymbol{\theta}' - \boldsymbol{\theta}'') \mathrm{d}^2\boldsymbol{\theta}' \mathrm{d}^2\boldsymbol{\theta}'' =$$

$$= \frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta}'') \int_{\mathbb{R}^2} U(|\boldsymbol{\theta}'|) \mathcal{D}^*(\boldsymbol{\theta}' - \boldsymbol{\theta}'' + \boldsymbol{\theta}) \mathrm{d}^2\boldsymbol{\theta}' \mathrm{d}^2\boldsymbol{\theta}'' =$$

$$= \frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta}'') \int_0^\infty U(x) \, x \int_0^{2\pi} \mathcal{D}^*(X - Y) \, \mathrm{d}\varphi \, \mathrm{d}x \, \mathrm{d}^2\boldsymbol{\theta}'' =$$

$$= -\frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta}'') \int_0^\infty U(x) \, x \int_0^{2\pi} \frac{1}{(xe^{i\varphi} - Y)^2} \, \mathrm{d}\varphi \, \mathrm{d}x \, \mathrm{d}^2\boldsymbol{\theta}'',$$

with the complex numbers $X = \theta_1' + i\,\theta_2' = xe^{i\varphi}$ and $Y = (\theta_1'' - \theta_1) + i\,(\theta_2'' - \theta_2)$. The complex integral is solved by using $\mathrm{d}\varphi = -i\,\dfrac{\mathrm{d}X}{X}$, transforming it into a loop integral and using the residual theorem:

$$\int_0^{2\pi} \frac{1}{(xe^{i\varphi} - Y)^2} \, \mathrm{d}\varphi = \frac{\pi}{Y^2} \left[ 2\,\Theta\left(|Y| - |X|\right) - |Y|\,\delta\left(|X| - |Y|\right) \right],$$

where $\Theta$ is Heaviside's step function and $\delta$ is Dirac's delta. The aperture mass becomes:

$$M_{ap}(\boldsymbol{\theta}) = -\frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta}'') \int_0^\infty \frac{U(x)x\pi}{Y^2} \left[ 2\Theta(|Y| - |X|) - |Y|\delta(|X| - |Y|) \right] \mathrm{d}x \, \mathrm{d}^2\boldsymbol{\theta}''$$

$$= \Re \int_{\mathbb{R}^2} \frac{\gamma(\boldsymbol{\theta}'')}{Y^2} \left[ |Y|^2 U(|Y|) - 2 \int_0^{|Y|} U(x) \, x \, \mathrm{d}x \right] \mathrm{d}^2\boldsymbol{\theta}'' =$$

$$= \Re \int_{\mathbb{R}^2} \frac{\gamma(\boldsymbol{\theta} + \boldsymbol{\theta}')}{(\theta')^2} \left[ |\boldsymbol{\theta}'|^2 U(|\boldsymbol{\theta}'|) - 2 \int_0^{|\boldsymbol{\theta}'|} U(x) \, x \, \mathrm{d}x \right] \mathrm{d}^2\boldsymbol{\theta}' =$$

$$= \int_{\mathbb{R}^2} \left[ -|\boldsymbol{\theta}'|^2 \Re\left( \frac{\gamma(\boldsymbol{\theta} + \boldsymbol{\theta}')}{(\theta')^2} \right) \right] \left[ \frac{2}{|\boldsymbol{\theta}'|^2} \int_0^{|\boldsymbol{\theta}'|} U(x) \, x \, \mathrm{d}x - U(|\boldsymbol{\theta}'|) \right] \mathrm{d}^2\boldsymbol{\theta}' =$$

$$= \int_{\mathbb{R}^2} \left[ -|\boldsymbol{\theta}' - \boldsymbol{\theta}|^2 \Re\left( \frac{\gamma(\boldsymbol{\theta}')}{(\theta' - \theta)^2} \right) \right] \left[ \int_0^{|\boldsymbol{\theta}' - \boldsymbol{\theta}|} \frac{2U(x)\,x}{|\boldsymbol{\theta}' - \boldsymbol{\theta}|^2} \, \mathrm{d}x - U(|\boldsymbol{\theta}' - \boldsymbol{\theta}|) \right] \mathrm{d}^2\boldsymbol{\theta}',$$

where we have introduced the complex numbers $\theta' = \theta_1' + i\,\theta_2'$ and $\theta = \theta_1 + i\,\theta_2$ corresponding to the vectors $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}$. The aperture mass can be written in a more immediate way if we define the *tangential shear component at position $\boldsymbol{\theta}'$ with respect to $\boldsymbol{\theta}$*:

$$\gamma_t(\boldsymbol{\theta}';\boldsymbol{\theta}) = -|\boldsymbol{\theta}' - \boldsymbol{\theta}|^2\,\Re\left(\frac{\gamma(\boldsymbol{\theta}')}{(\theta' - \theta)^2}\right) = -\Re\left(\frac{\gamma(\boldsymbol{\theta}')\,(\theta' - \theta)^*}{(\theta' - \theta)}\right), \qquad (2.62)$$

and a *filter function*, a real function of a real variable $t$:

$$Q(t) = \frac{2}{t^2}\int_0^t U(t')\,t'\,\mathrm{d}t' - U(t). \qquad (2.63)$$

The aperture mass can thus be written as:

$$M_{ap}(\boldsymbol{\theta}) = \int_{\mathbb{R}^2}\gamma_t(\boldsymbol{\theta}';\boldsymbol{\theta})\,Q(|\boldsymbol{\theta}' - \boldsymbol{\theta}|)\,\mathrm{d}^2\boldsymbol{\theta}'. \qquad (2.64)$$

This expression allows us to express the aperture mass, a $U$-weighted integral of the convergence, as an integral over the tangential shear weighted by the filter function $Q$. Different filters have different properties; generally they have a compact support, so that $M_{ap}(\boldsymbol{\theta})$ can be expressed as an integral of the tangential shear over a finite area around $\boldsymbol{\theta}$, and with a suppression in the aperture center so that the tangential shear does not diverge and the weak lensing approximation ($\kappa \ll 1$, $|\gamma| \ll 1$) does not break down. It is generally possible to create signal-to-noise ratio map of the mass aperture statistic with a chosen filter $Q$, in which sufficiently strong lenses will appear as significant peaks. This method has been tested on synthetically generated data, and it is shown in several works that it lives up to the expectations (Maturi et al. 2005; Pace et al. 2007). The signal-to-noise ratio map is not corrupted by small-scale deflectors (such as individual galaxies), nor by larger-scale deflectors. This method can then be used to search for mass concentrations (dark haloes of galaxy clusters) given wide-field images of ellipticities.

### 2.2.1   On the relation between aperture mass and shear

We have defined the tangential shear in equation (2.62) as the real part of a given complex number. If we consider the imaginary part we can define the equivalent *radial* or *cross shear component at position $\boldsymbol{\theta}'$ with respect to $\boldsymbol{\theta}$*:

$$\gamma_r(\boldsymbol{\theta}';\boldsymbol{\theta}) = -|\boldsymbol{\theta}' - \boldsymbol{\theta}|^2\,\Im\left(\frac{\gamma(\boldsymbol{\theta}')}{(\theta' - \theta)^2}\right) = -\Im\left(\frac{\gamma(\boldsymbol{\theta}')\,(\theta' - \theta)^*}{(\theta' - \theta)}\right). \qquad (2.65)$$

The tangential and radial shear components can also be expressed as explicit functions of the cartesian components of $\Delta\theta = \theta' - \theta = \Delta\theta_1 + i\,\Delta\theta_2$ as:

$$\gamma_t(\boldsymbol{\theta}';\Delta\theta) = -\frac{1}{\Delta\theta_1^2 + \Delta\theta_2^2}\left[\gamma_1(\boldsymbol{\theta}')\,(\Delta\theta_1^2 - \Delta\theta_2^2) + 2\gamma_2(\boldsymbol{\theta}')\,(\Delta\theta_1\,\Delta\theta_2)\right] \qquad (2.66)$$

$$\gamma_r(\boldsymbol{\theta}';\Delta\theta) = -\frac{1}{\Delta\theta_1^2 + \Delta\theta_2^2}\left[\gamma_2(\boldsymbol{\theta}')\,(\Delta\theta_1^2 - \Delta\theta_2^2) - 2\gamma_1(\boldsymbol{\theta}')\,(\Delta\theta_1\,\Delta\theta_2)\right] \qquad (2.67)$$

Let us note that the cross shear is equivalent to computing the tangential shear on a map where each galaxy shear has been rotated by $\pi/4$, that according to (2.38) means using a new shear $\gamma'$ whose components are:

$$\gamma'_1 = \gamma_2, \quad \gamma'_2 = -\gamma_1. \tag{2.68}$$

The tangential and radial shear can also be expressed as a function of the radial coordinates of $\Delta\theta = \vartheta\,(\cos\varphi + i\,\sin\varphi)$ as:

$$\gamma_t(\boldsymbol{\theta'};\varphi) = -\gamma_1(\boldsymbol{\theta'})\cos(2\varphi) - \gamma_2(\boldsymbol{\theta'})\sin(2\varphi) \tag{2.69}$$

$$\gamma_r(\boldsymbol{\theta'};\varphi) = +\gamma_1(\boldsymbol{\theta'})\sin(2\varphi) - \gamma_2(\boldsymbol{\theta'})\cos(2\varphi) \tag{2.70}$$

These quantities are directly observable in the weak-lensing limit where $\kappa \ll 1$, $\gamma \ll 1$. Let us fix the reference point of the tangential and cross components $\boldsymbol{\theta}$ and let us compute their average value on a circle around $\boldsymbol{\theta}$ that has a fixed radius $\vartheta$. Using the two-dimensional version of Gauss' theorem, it is possible to show:

$$\langle\gamma_t\rangle(\vartheta;\boldsymbol{\theta}) = \frac{1}{2\pi}\oint \gamma_t(\theta + \vartheta\,e^{i\,\varphi})\,\mathrm{d}\varphi = \frac{\bar{\Sigma}-\Sigma}{\Sigma_{crit}} = \frac{\Delta\Sigma}{\Sigma_{crit}}(\vartheta;\boldsymbol{\theta}) \tag{2.71}$$

$$\langle\gamma_r\rangle(\vartheta;\boldsymbol{\theta}) = \frac{1}{2\pi}\oint \gamma_r(\theta + \vartheta\,e^{i\,\varphi})\,\mathrm{d}\varphi = 0 \tag{2.72}$$

where $\Delta\Sigma(\vartheta;\boldsymbol{\theta})$ is the excess surface mass density around $\boldsymbol{\theta}$; it is the azimuthally averaged surface density profile in a circle of radius $\vartheta$ around $\boldsymbol{\theta}$ (Umetsu 2020). Equation (2.72) shows that the cross shear component averaged around the loop extract the B-mode distortion pattern, since it is expected to be statistically consistent with zero if the signal is due to weak lensing. Therefore, a measurement of the B-mode signal provides a useful null test against systematic errors. Equation (2.71) shows that, given an arbitrary circular loop of radius around the chosen centre $\boldsymbol{\theta}$, the tangential shear component averaged around the loop extracts a signal that is only due to the convergence; it thus extracts the E-mode distortion pattern (see section 2.1.4). Another fact worth mentioning is that if one rotates its shear map by $\pi/4$ the E-mode and B-mode are swapped.

## 2.2.2 On observable galaxy ellipticities

There is a link between the reduced shear $g$, the intrinsic complex ellipticity $\epsilon^{(s)}$ of a source and the complex ellipticity of its image $\epsilon$. For an elliptical source with axis ratio $r \leq 1$ the complex ellipticity has modulus $|\epsilon^{(s)}| = \frac{|1-r|}{|1+r|}$ and phase given by twice the angle between the major axis and the positive $\theta_1$-direction. For sub-critical lenses:

$$\epsilon = \frac{\epsilon^{(s)} + g}{1 + g^*\,\epsilon^{(s)}}. \tag{2.73}$$

Assuming that the intrinsic orientation of the sources is random, the expectation value of the average ellipticity of a local sample of galaxy images becomes

$$\langle \epsilon \rangle = g \approx \gamma, \tag{2.74}$$

where $g \approx \gamma$ holds in the weak lensing approximation $\kappa \ll 1$. Thus, the ellipticity of a galaxy image is an unbiased estimate of the local shear in the case of weak lensing.

Let us suppose we have a dataset of complex galaxy ellipticities $\epsilon_i$ in position $\boldsymbol{\theta_i}$, with a number density of galaxy images $n_g$. We can define the observational version of equation (2.62) as the *complex tangential ellipticity of the image in position $\boldsymbol{\theta_i}$ with respect to $\boldsymbol{\theta}$* as

$$\epsilon_t(\boldsymbol{\theta_i}; \boldsymbol{\theta}) = \epsilon_{t\,i}(\boldsymbol{\theta}) = -\Re\left(\frac{\epsilon_i \left(\theta_i - \theta\right)^*}{\left(\theta_i - \theta\right)}\right) \tag{2.75}$$

and the observational estimator of the mass aperture (2.64) as:

$$M_{ap}(\boldsymbol{\theta}) = \frac{1}{n_g} \sum_i \epsilon_{t\,i}(\boldsymbol{\theta})\, Q(|\boldsymbol{\theta_i} - \boldsymbol{\theta}|), \tag{2.76}$$

where the sum extends over all galaxy images within the aperture, i.e., within the support of the filter $Q$ (where $Q \neq 0$). It can be shown that the variance of this estimator is given by:

$$\sigma^2_{M_{ap}}(\boldsymbol{\theta}) = \frac{1}{2n_g} \sum_i |\epsilon_{t\,i}(\boldsymbol{\theta})|^2\, Q^2(|\boldsymbol{\theta_i} - \boldsymbol{\theta}|). \tag{2.77}$$

The noise produced by the intrinsic galaxy ellipticities can be taken as the square root of this variance, and the signal-to-noise ratio can then be computed; it is independent of the normalization of the weight function.

# Chapter 3

# The implementation of an optimal filter: Get the Halo 2021

The main aim of this thesis is to construct a `C++` code that applies an "optimal" linear matched filter to weak-lensing data in order to detect dark-matter haloes corresponding to galaxy clusters. The filter was first developed by Maturi et al. (2005) where it was applied on simulated data, and then applied on real data in Maturi et al. (2007). The "matched filter" technique is a method (Schneider 2015) to build catalogs of clusters of galaxies based on the assumption that clusters are not just a collection of galaxies, but that their galaxy overdensities possess certain given properties. For example one could assume that the luminosity distribution of the galaxies follows a Schechter-type luminosity function if optical data is available. The filter here implemented assumes a mean radial profile of the halo shear pattern and a power spectrum for the noise; it is defined in the Fourier domain as proportional to their ratio, so that the modes that can attributed to lensing by the halo profile are amplified and the modes due to the noise are suppressed. In order to separate dark matter haloes from spurious peaks caused by the large scale structure the noise is composed by three contributions: Poisson noise from galaxy counts, noise from intrinsic ellipticities and noise related to the weak lensing by large scale structure. The assumptions of the optimal filter implemented will be explained in section 3.1. Section 3.2 will describe the numerical code that implements the filter, *Get the Halo* 2021. This program is a new version of the original *Get the Halo* realised by Matteo Maturi more than fifteen years ago and used in Maturi et al. (2005) and Maturi et al. (2007). The main perks of this new version are that it is compatible with the 2017 standard of the `C++` language and it replaces the *astrocpp* library with the *CosmoBolognalib* (Marulli et al. 2016).

The *CosmoBolognaLib* are a large set of Open Source `C++` numerical libraries for cosmological calculations, although they can also be used in other high-level scriptinh languages through wrapping, for instance `Python` codes. It is a *living* project aimed at defining a common numerical environment for cosmological investigations of the

large-scale structure of the Universe. One of its primary focuses is to help in handling astronomical catalogues, both real and simulated, measuring one-point, two-point and three-point statistics in configuration space, performing cosmological analyses and deriving cosmological constraints. They can be freely downloaded at the *CosmoBolognalib* website[1].

## 3.1 On the theory of the optimal filter

The dark matter distribution in the Universe is not a discrete distribution that contains only haloes, but it can be thought as a continuous distribution where the haloes are peaks of mass density embedded into a lower density *large-scale structure*, the cosmic web that connects them. When someone measures the inhomogeneities of projected mass distribution through lensing observations, the signal of haloes is overlaid by the lensing signal of the large-scale structure in front of and behind the haloes. Being approximately a Gaussian random field, lensing by large-scale structure adds peaks and troughs to the signal which can be mistaken for haloes, so part of the claimed dark-halo detections can actually be peaks in the random weak gravitational lensing signal of the large-scale structure. It is not possible to strictly separate these two types of signal because of the unsharp boundary between the haloes and the large-scale structure, but it is possible to define a typical scale that helps discriminate between their lensing signals. The large-scale structure can be considered as composed by dark-matter haloes of a broad and continuous mass range. At each cosmological epoch there is a characteristic scale length, the *nonlinear scale*, at which the variance of the dark matter density contrast becomes unity. This was the scale described throughout chapter 1 and given in equation (1.51). It separates the small-scale regime where the dark matter power spectrum is dominated by the contributions of presumably virialised haloes from the large-scale regime where the dark-matter density can be considered as a linear superposition of linearly evolved perturbation modes.

This scale suggests the operational definition of the lensing signal due to the haloes as the signal contributed by non-linear, gravitationally bound, virialised structures; the lensing signal due to the large-scale structure is instead the signal contributed by the linearly evolved matter distribution beyond the non-linear scale. It was first shown in Maturi et al. (2005) and Maturi et al. (2007) that these definitions can be used to construct a linear matched filter capable to detect the weak lensing signal of dark matter haloes corresponding to galaxy clusters while also suppressing the signal of the large-scale structure, thus reducing the spurious detections attributed to random peaks in the projected mass density of the large-scale structure. The underlying assumption of this approach is that the filter is searching for those haloes that do create the non-linear power spectrum, while the linearly evolved structures are treated as a noise component

---

[1] https://github.com/federicomarulli/CosmoBolognaLib

that modifies the shape of the matched filter. The filter can also be chosen to maximise the signal-to-noise ratio given a mass profile of the cluster halo. In particular we want to choose an optimal filter that maximise the S/N ratio given a Navarro-Frenk-White (NFW) profile and given that the noise is made up by the three above mentioned contributions: shot noise from galaxy counts; white noise from intrinsic ellipticities of galaxies; noise from the lensing effects produced by the random projected peaks of the large-scale structure.

Let us consider the weak gravitational lensing signal of dark matter halo as a function $S(\boldsymbol{\theta})$ that can be factorised into its amplitude $A$ and angular shape $\tau(\boldsymbol{\theta})$; the data we measure $D(\boldsymbol{\theta})$ are contaminated by some noise $N(\boldsymbol{\theta})$, so they can be written as:

$$D(\boldsymbol{\theta}) = S(\boldsymbol{\theta}) + N(\boldsymbol{\theta}) = A\,\tau(\boldsymbol{\theta}) + N(\boldsymbol{\theta}). \tag{3.1}$$

In this application the signal will be the shear tangential components. The noise is made up by several contributions, that we are going to describe in the following section 3.1.1.

### 3.1.1   On noise properties

The noise contributions are conveniently described in the Fourier space, where their correlation functions allow us to define the *noise power spectrum $P_N(k)$*:

$$\langle \tilde{N}(\boldsymbol{k})\,\tilde{N}^*(\boldsymbol{k}')\rangle = (2\pi)^2\,\delta_D^{(2)}(\boldsymbol{k}' - \boldsymbol{k})\,P_N(k). \tag{3.2}$$

It is possible to identify three main noise components:

- Poisson shot noise proportional to the number density of galaxies $n_g$. This noise arises because the galaxy ellipticities are measured at random positions. It is constant in the Fourier domain (white noise).

- White noise due to the intrinsic ellipticities of the galaxies. This noise arises because the determination of a single galaxy ellipticity is a very noisy measurement of the shear; the added noise is proportional to the variance of the intrinsic galaxy ellipticity $\sigma_{\epsilon_s}^2$ and constant in the Fourier domain.

- Noise caused by the large-scale structure. When the matter distribution of the large-scale structure is projected along the line of sight random peaks made up by uncorrelated mass might arise. They contaminate the signal by adding spurious peaks in the signal, unless they are considered as noise. This is done by modelling this signal using the linear dark-matter power spectrum, according to the operational separation between halo and large-scale structure lensing given in section 3.1.

These noise components are assumed to be random with zero mean and isotropic such that their statistical properties are independent of the position on the sky. These assumptions are well justified, since in first approximation the background galaxies are randomly

positioned and oriented, and weak lensing by the large-scale structures is well described by an isotropic Gaussian random field. Thus the noise given by the intrinsic ellipticity of the sources combined with their finite number is modelled by its power spectrum $P_\epsilon$ (constant), and the noise caused by weak lensing of intervening large-scale structures is modelled by the cosmic-shear power spectrum $P_\gamma(k)$ derived from the linear dark-matter power spectrum. The complete noise power spectrum is thus:

$$P_N(k) = P_\epsilon + P_\gamma(k) = \frac{1}{2}\frac{\sigma_{\epsilon_s}^2}{n_g} + \frac{1}{2}P_\kappa(k). \tag{3.3}$$

The cosmic shear power spectrum is equal to the power spectrum of the convergence $P_\kappa$ generated by the large scale structure, and the factor $1/2$ arises because only one component of the ellipticity (the tangential shear) contributes to the measurement, thus $P_\gamma = \frac{1}{2}P_\kappa$. This power spectrum is (Umetsu 2020):

$$P_\kappa(k) = \frac{9H_0^4\,\Omega_m^2}{4c^4}\int_0^{\chi_s}\frac{\bar{W}(\chi;\chi_s)}{a^2(\chi)}P_\delta\left(\frac{k}{f_K(\chi)};\chi_s\right)\mathrm{d}\chi \tag{3.4}$$

where $\chi$ is the comoving distance, $\chi_s$ the comoving distance of the sources (for the sake of simplicity we assume that all our sources have the same comoving distance from the observer, and the same redshift), $a$ is the scale factor of the Universe, $\bar{W}$ is a weight function, $P_\delta$ is the power spectrum of the dark matter tridimensional density fluctuations, $K$ is the spatial curvature, $f_K(\chi)$ is the comoving angular diameter distance described in (1.6). The weight function $\bar{W}$ is given, under the approximation of having all sources at the same redshift by:

$$\bar{W}(\chi;\chi_s) = \frac{f_K(\chi_s - \chi)}{f_K(\chi_s)}. \tag{3.5}$$

This power spectrum can be computed thanks to the *CosmoBolognalib* if a source redshift is given; its value represents a typical redshift of the sources, for instance the mean or the median redshift of the source galaxies. The `C++` code that computes the cosmic shear power spectrum inside the *CosmoBolognalib* uses the function *cbl::cosmology::Cosmology::Pk_DM* to compute $P_\delta$, and it assumes a $\Lambda$CDM cosmology with the default cosmological parameters available in the *CosmoBolognalib*. The main cosmological parameters are the normalised Hubble constant $h = 0.7$ and the density parameters of the matter $\Omega_{m,0} = 0.27$, of the cosmological constant $\Omega_{\Lambda,0} = 0.73$, of the baryons $\Omega_{bar,0} = 0.046$ and of the radiation $\Omega_{rad,0} = 0.0$.

### 3.1.2 On the model of the halo lensing signal

Theoretical assumptions on the spatial distribution of the signal are required to build the filter. The first assumption is that galaxy clusters are on average axially symmetric,

thus $\tau(\boldsymbol{\theta}) = \tau(|\boldsymbol{\theta}|)$. The second is that the cluster halo (the lens) possesses a Navarro-Frenk-White (NFW) density profile, since several works have shown that it is a good analytic model to describe the density profile of the haloes (J. F. Navarro et al. 2004; Julio F. Navarro et al. 1996). The profile is

$$\rho(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)\left(1 + \frac{r}{r_s}\right)^2}, \tag{3.6}$$

where $r_s$ represents the characteristic scale radius, where the logarithmic slope $\dfrac{\mathrm{d}\ln\rho}{\mathrm{d}\ln r}$ equals $-2$. The normalization $\rho_s = 4\rho(r_s)$ is a characteristic density. When $r \ll r_s$ the density decreases as $\rho \propto r^{-1}$, when $r \gg r_s$ as $\rho \propto r^{-3}$. The two parameters of the NFW profile are not independent, this profile actually depends only on one parameter. The *overdensity mass* $M_\Delta$ is given by integrating equation (3.6) up to the *overdensity radius* $r_\Delta$ at which the mean internal density is $\Delta$ times the critical density of the Universe at halo redshift $\rho_c(z_h)$ as defined in equation (1.21). For a NFW halo the *concentration parameter* can be defined as $c_\Delta \equiv \dfrac{r_\Delta}{r_s}$; the typical density is then:

$$\rho_s = \frac{\Delta}{3} \frac{c_\Delta^3}{\ln(1 + c_\Delta) - \frac{c_\Delta}{1+c_\Delta}} \rho_c(z_h). \tag{3.7}$$

The NFW profile thus depends only on the concentration parameter $c_\Delta$; $\Delta$ is typically chosen to be 200. The gravitational lensing properties of the NFW lens have been widely explored (Meneghetti et al. 2003). Its lensing potential is:

$$\psi(x) = 4\kappa_s h(x), \tag{3.8}$$

with $x = \frac{r}{r_s}$ (projected distance from the lens center in units of the distance scale),

$$h(x) = \frac{1}{2}\ln^2\frac{x}{2} + \begin{cases} -2\operatorname{arctanh}^2\sqrt{\frac{1-x}{1+x}} & (x < 1) \\ 0 & (x = 1) \\ 2\arctan^2\sqrt{\frac{x-1}{x+1}} & (x > 1) \end{cases} \tag{3.9}$$

and $\kappa_s = \rho_s\, r_s\, \Sigma_{crit}^{-1}$. The shear profile can be computed from the potential by using equations (2.25) and (2.26), the convergence using equation (2.22) and the reduced shear profile with equations (2.41) (Wright et al. 2000). The Fourier transform of the reduced shear profile $g(\boldsymbol{x})$

$$\tilde{\tau}(\boldsymbol{k}) = \tilde{g}(\boldsymbol{k}) = \int_{\mathbb{R}^2} g(\boldsymbol{x})\, e^{i\,\boldsymbol{x}\cdot\boldsymbol{k}}\, \mathrm{d}^2\boldsymbol{x} \tag{3.10}$$

is the halo signal necessary to compute the filter. The reduced shear of a NFW halo is shown in figure 3.1, where panel 3.1a shows $g(|\boldsymbol{x}|)$ in the real space and panel 3.1b shows $\tilde{g}(|\boldsymbol{k}|)$ in the Fourier space.

**(a)** Real space
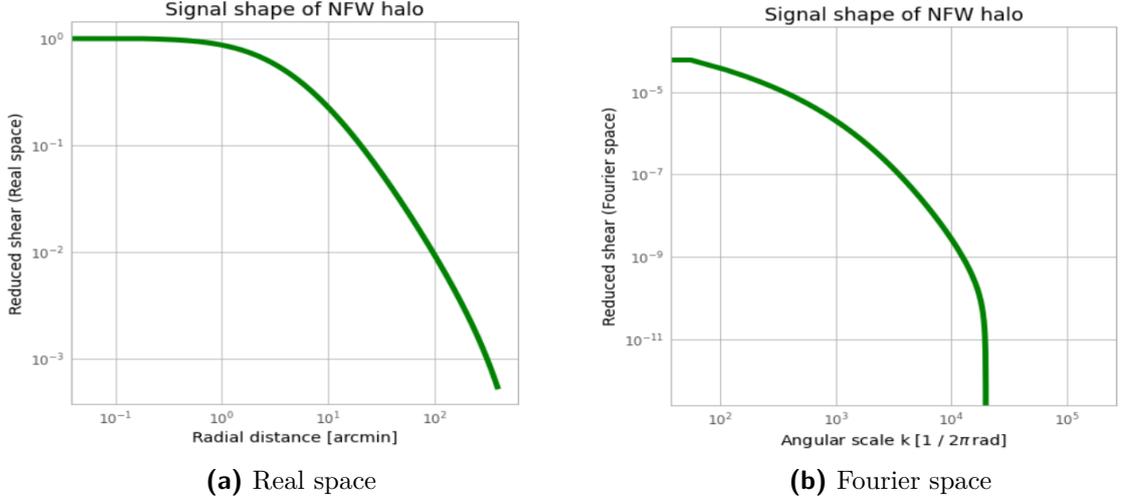
**(b)** Fourier space

**Figure 3.1:** Reduced shear profile from a NFW halo in the real (**panel 3.1a**) and Fourier space (**panel 3.1b**). The halo used to compute them has mass $10^{15}\,\mathrm{M}_\odot\,h^{-1}$, it is located at redshift 0.3 with background sources at redshift 1.0. Its scale radius projected on the sky is $4'$.

### 3.1.3 On the definition of the optimal filter

Once the noise power spectrum and the expected halo signal have been computed they can be used to define the filter. The linear filter $\Psi$ is a function that, when convoluted with the data $D(\boldsymbol{\theta})$ as defined in equation (3.1), yields an estimate of the amplitude of the signal. We define the estimator for this amplitude as in Maturi et al. (2005):

$$A_{\mathrm{est}}(\boldsymbol{\theta}) = \int_{\mathbb{R}^2} D(\boldsymbol{\theta}')\Psi(\boldsymbol{\theta} - \boldsymbol{\theta}')\mathrm{d}^2\boldsymbol{\theta}'. \tag{3.11}$$

The estimator must satisfy two constraints. First, it must be unbiased: its average error $b$ has to vanish ($b = 0$):

$$b \equiv \langle A_{\mathrm{est}} - A \rangle = A \left[ \int_{\mathbb{R}^2} \tau(\boldsymbol{\theta}')\Psi(\boldsymbol{\theta}')\mathrm{d}^2\boldsymbol{\theta}' \right]. \tag{3.12}$$

Second, the measurement noise $\sigma$, determined by the mean-squared deviation of the estimate from its true value,

$$\sigma^2 = \langle (A_{\mathrm{est}} - A)^2 \rangle = b^2 + \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} |\tilde{\Psi}(\boldsymbol{k})|^2 P_N(k)\,\mathrm{d}^2\boldsymbol{k}$$

$$= b^2 + \frac{1}{2\pi} \int_0^\infty |\tilde{\Psi}(\boldsymbol{k})|^2 P_N(k) k\,\mathrm{d}k \tag{3.13}$$

has to be minimal. To find a filter $\Psi$ satisfying these two conditions, we combine them by means of a Lagrangian multiplier $\lambda$, carry out the variation of $\mathscr{L} = \sigma^2 + \lambda b$ with

respect to $\Psi$ and thus find the function $\Psi$ minimizing $\mathscr{L}$. The solution of this variational minimization in the Fourier domain is given by:

$$\tilde{\Psi}(\boldsymbol{k}) = \frac{1}{(2\pi)^2} \left[ \int_{\mathbb{R}^2} \frac{|\tilde{\tau}(\boldsymbol{k})|^2}{P_N(k)} \, \mathrm{d}^2\boldsymbol{k} \right]^{-1} \frac{|\tilde{\tau}(\boldsymbol{k})|^2}{P_N(k)}, \tag{3.14}$$

where $\tau$ is the expected shear profile of a NFW halo and $P_N$ is the noise power spectrum, thus the filter is constructed to be most sensitive for those spatial frequencies where the signal $\tilde{\tau}$ is large and the noise $P_N$ is low. The convolution to compute the estimator is done in the real domain, so an anti-Fourier transform must be applied to $\tilde{\Psi}(k)$ to obtain $\Psi(\theta)$.

Typical trends of the optimal filter are shown in figure 3.2 for both the real and Fourier space. In the Fourier space the filter typically shows a peak at those frequencies with a high ratio between the reduced shear of the NFW halo $\tilde{\tau}$ and the noise power spectrum $P_N$. In the real space it shows a central peak, responsible for creating signal spots in the output maps where there are haloes, and a negative region outside the peak, responsible for creating troughs with lower signal outside the spots. When the signal processed by the filter is the tangential shear, then a typical signature of a halo detection on the output maps is a spot with high signal-to-noise ratio (S/N) surrounded by a trough with lower S/N. These structures can be seen in figure 3.3, where we show a typical output map processed with this filter.

When the signal is the tangential shear the estimator (3.11) is mathematically similar to an Aperture mass as defined in equation (2.64):

$$A_{\mathrm{est}}(\boldsymbol{\theta}) = \int_{\mathbb{R}^2} \gamma_t(\boldsymbol{\theta}'; \boldsymbol{\theta}) \Psi(|\boldsymbol{\theta}' - \boldsymbol{\theta}|) \mathrm{d}^2\boldsymbol{\theta}'. \tag{3.15}$$

Their difference is that the optimal weak lensing filter is not defined from an arbitrary compensated weight function according to equation (2.63), but defined so that it maximises the signal-to-noise ratio and minimises the contamination induced by the large-scale structure; its signal-to-noise ratio properties differs from that of the aperture mass.
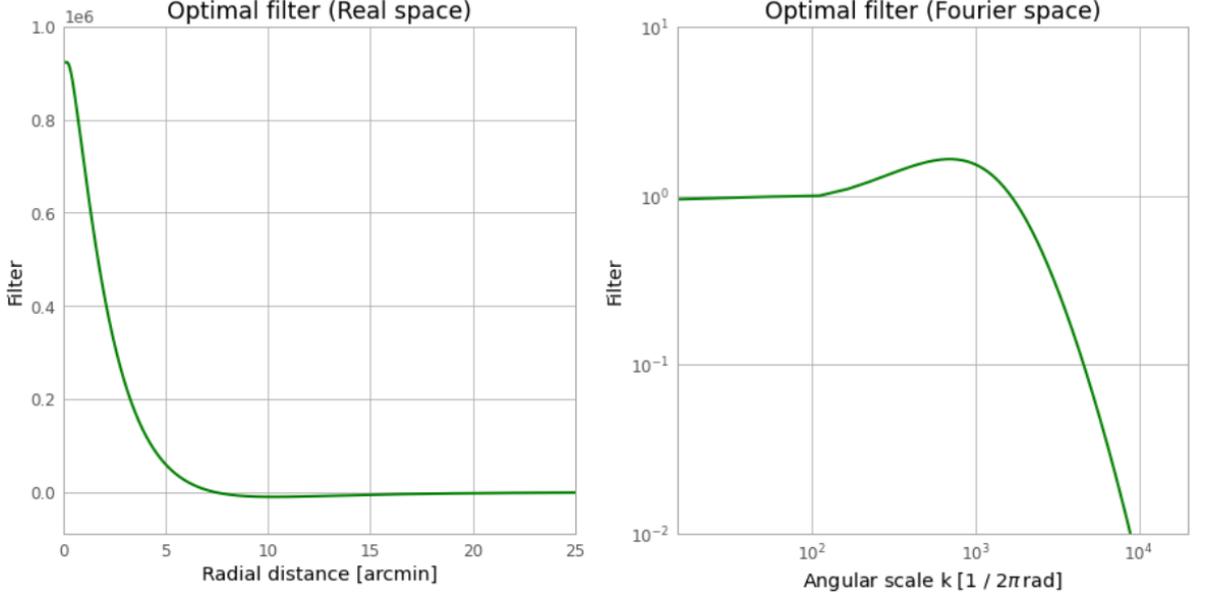
**Figure 3.2:** We show the optimal filter in the real domain, $\Psi$ on the left, and in the Fourier space, $\tilde{\Psi}$ on the right. This filter was built by taking as a model a NFW halo of virial mass $10^{15}\,\mathrm{M}_\odot\,h^{-1}$ at redshift 0.3. The scale radius of the NFW halo model projected on the sky is $4'$. The background sources are at redshift 1.0, used to compute the noise power spectrum of the large scale structure $P_\gamma$. The galaxy white noise is $P_\epsilon = 1.28 \times 10^{-10}$. These quantities allow us to compute the filter in the Fourier space $\tilde{\Psi}$ as described in equation (3.14). A Fourier transform allows us to get the filter $\Psi$ in the real space. It has a typical scale (average radius) of $2.2'$. In the **right panel** we can see that the filter has a peak around those spatial frequencies with a high ratio between the reduced shear of the NFW halo $\tilde{\tau}$ and the noise power spectrum $P_N$ as described in equation (3.14). In the **left panel** we can see the filter in the real space. It has a central peak, responsible for creating signal spots in the output maps where there are haloes, and a negative region outside the peak, responsible for creating troughs with lower signal outside the spots. These structures can be seen in figure 3.3, where we show an output map of tangential shear convoluted with this filter according to equation (3.11). A typical signature of a halo detection on the output map is a spot with high signal-to-noise ratio (S/N) surrounded by a trough with lower S/N.

## 3.2 On Get the Halo 2021

*Get the Halo* 2021 (GtH21 for short) is a `C++` program that, given a dataset of galaxy positions and shear and the optimal linear matched filter for that dataset (they must be prepared in appropriate input files) computes the weak lensing signal estimator (3.11) and other quantities described in section 3.2.2. The output file is then analysed as described in section 3.2.4 to get a catalogue of detected haloes.

### 3.2.1 On the necessary input files

The program requires three input files.

**Weak lensing data file.** The first input file necessary to execute GtH21 is an ASCII file with the weak lensing data. This file must be structured as a table where each row is an observed galaxy and each column is one of its measured quantities. There are four mandatory columns: the two sky coordinates $x$ and $y$ of the galaxy, the two components of its ellipticity $g_1$ and $g_2$ (expected values for the reduced shear). Then there are three optional columns: the redshift of the galaxy $z_s$, the weak lensing bias factor $m$ and the weak lensing weight $w$ of the galaxy. The galaxies' redshifts are used only to compute the average source redshift; it is requested to compute the cosmic shear power spectrum given in equation (3.4) that must be included as a source of noise according to equation (3.13). If the redshift column is not given GtH21 will just require the user to insert a typical source redshift from the terminal. The bias factor $m$ is used to take into account several observational and instrumental effects that affect the ellipticities measurements when galaxies are observed. Some of them are described by Umetsu (2020): the impact of noise on the galaxy shape measurement (both statistical and systematic uncertainties), the isotropic smearing component of the PSF and the effect of instrumental PSF anisotropy. The calibration of the signal response is considered through the bias factor

$$g_{true} = (1 + m)\, g_{obs}. \tag{3.16}$$

The weights are generally the inverse-variance weights of the ellipticity. The galaxy weights that GtH21 assigns to each galaxy in the ellipticity map are normalised to their sum. The first input file must be prepared by converting the requested data, in general given as a catalogue of galaxy ellipticities, into an ASCII file.

**Optimal filter files.** The second input file necessary to execute GtH21 is an ASCII file structured as a table that contains the optimal filter $\Psi$ in the real domain. A column of this file must be the radial positions $r$ in *arcminutes* at which the filter is evaluated. Another column of the file must be the value of the filter $\Psi(r)$ evaluated at the radial position $r$ contained in the same row. Those two columns are mandatory. The radial

values must be ordered for ascending values and the $r$ values must be equi-spaced: the filter must be sampled regularly and linearly. This condition is necessary for a correct linear interpolation of the filter function. It is then possible to add a third optional column that contains the values of the reduced shear of the chosen NFW halo $g(r)$ evaluated at the radial position $r$. If this column is included GtH21 will also compute a local normalization map for the weak lensing estimator (3.13), as it will be described in equation (3.21).

In order to compute the correct noise given in equation (3.13) the code needs the filter in the Fourier domain $\tilde{\Psi}(k)$. It can be given in a third ASCII file where the first column is the angular mode $k$ and the second column is the filter $\tilde{\Psi}$ evaluated at the $k$ contained in the same row.

The second and third input file must be produced separately with a program that implements equation (3.14) and its Fourier transform. We have used a specific program to produce the weak lensing filter. This program requires the following input parameters:

- **Lens mass**. This is the mass of the NFW dark matter halo that acts as the typical lens, in units of $M_\odot\, h^{-1}$. In Maturi et al. (2007) it is shown that the detected haloes are not very sensitive on this parameter.

- **Lens redshift**. This is the redshift of the typical halo we want to detect.

- **Average source redshift**. The lens and source redshift are needed to set physical quantities of any lensing system, such as the critical surface density defined in equation (2.12).

- **White noise amplitude**. This is the noise contribution of the power spectrum given by the galaxies, described as $P_\epsilon$ in section 3.1.1.

- **Instrumental beam FWHM**. This is the resolution with which the filter is defined. It should be approximately equal to the average galaxy separation, computed as the inverse of the average surface number density of galaxies. It must be given in *arcminutes.*

- **Power noise spectrum file**. File that contains the cosmic shear power spectrum of the convergence $P_\gamma$.

- File with the **cosmological parameters**. Optional parameter, it allows one to change the default cosmological parameters.

- **Scale radius** in arcminutes. Optional parameter that allows to set the halo dimension not through the mass but through the projected scale radius.

**Setup file.** All these input files are given to GtH21 through a single *setup file.* The program must be executed from the terminal by giving the executable one argument that contains the name of the setup file. It is a text file where each line is structured as:

$$\text{PARAMETER} = \text{VALUE} \ \# \ \text{Optional comment} \tag{3.17}$$

The parameters are read by GtH21 thanks to a class "initializer" we wrote to read data formatted like in example (3.17). If $\#$ appears in a line the rest of the line is considered a comment. A typical setup file contains several parameters describing the directory and names of the input files, how they are structured, how to define the grid for the output maps of GtH21 and other useful parameters. We report them all in appendix A.

Once the data file, filter files and the initialization file have all been prepared GtH21 can be executed from the terminal. The name of the executable must be followed by one additional argument, the name of the initialization file. If it is missing or there are more names than the program just prints some instructions on how it can be properly used.

### 3.2.2 On the output file written by Get the Halo 2021

If the program is executed correctly it produces a single FITS file with seven image frames contained inside *Header-Data-Units* (HDU).

The first or *primary HDU* contains the weak lensing signal; it computes the estimator (3.11) by using the tangential shear as data.

$$E_{map}(\boldsymbol{\theta}) \equiv A_{est}(\boldsymbol{\theta}) = \frac{1}{n_g} \sum_k \gamma_t(\boldsymbol{\theta}; \boldsymbol{\theta_k}) \, w_k \, \Psi(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|) \tag{3.18}$$

where the sum is extended to all the galaxies with position $\boldsymbol{\theta_k}$ less distant from the map point $\boldsymbol{\theta}$ than an appropriate *cut radius*, the angular distance for which the filter $\Psi$ is consistent with 0. If the lensing weights $w_k$ are not used they are all set equal to 1. The normalization factor $n_g$ is the effective number density of galaxies. This is the number density of perfectly measured galaxies that would contribute the same amount of shear noise as the (imperfectly) measured ensemble of galaxies; it was introduced (Albrecht et al. 2006) when considering the Fisher matrix for a weak lensing survey, and it is computationally found by summing the lensing statistical weights divided by their maximum. Because of its properties the effective number density of galaxies should be used instead of the ordinary number density when a statistics requires it (Chang et al. 2013). This is done in order to enhance the reliability of the results from a weak lensing survey, since their reliability depends on the total sky coverage and the number density of galaxies with accurate shear measurements.

The tangential shear is computed by using equation (2.66). Equation (3.18) implements the aperture mass estimator defined in (2.76) when the filter function is the optimal

linear filter in the real domain, defined by the Fourier transform of (3.14). In section 2.2.1 it was shown that tangential shear extracts the E-mode of the lensing shear, so the weak lensing signal map $A_{est}(\boldsymbol{\theta})$ is also called *E-mode map*.

The second HDU in the output file is the *B-mode map*:

$$B_{map}(\boldsymbol{\theta}) = \frac{1}{n_g} \sum_k \gamma_r(\boldsymbol{\theta}; \boldsymbol{\theta_k}) \, w_k \, \Psi(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|). \tag{3.19}$$

It is defined exactly as equation (3.18), the only difference is that is uses the radial shear defined in equation (2.67), so it extracts the B-mode of lensing shear, that is expected to be consistent with 0. This map is then used to check for systematic errors.

The third HDU is a map of the shear modulus variance:

$$\sigma_g^2(\boldsymbol{\theta}) = \frac{1}{2\,n_g^2} \sum_k |g(\boldsymbol{\theta_k})|^2 \, w_k^2 \, \Psi^2(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|), \tag{3.20}$$

that can be used as an alternative noise estimator when the local shear is comparable to the intrinsic galaxy ellipticities (Maturi et al. 2007). It is not applied in the work of this thesis.

The fourth HDU is the *galaxy number density map* on the output map.

The fifth HDU is a *local normalization map*, computed only if the expected weak lensing signal from a halo is given to the program, that is the radial reduced shear profile $\tau(\vartheta)$ with $\vartheta$ angular radial distance from the halo center. The map is computed as:

$$A_{norm}(\boldsymbol{\theta}) = \frac{1}{n_g} \sum_k \tau(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|) \, w_k \, \Psi(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|). \tag{3.21}$$

To measure the correct signal and noise our estimators should be multiplied for this normalization, but they would cancel out when computing the signal-to-noise ratio.

The sixth HDU is the noise estimate for the variance obtained as $\sigma^2 = \langle (A_{est} - A)^2 \rangle$:

$$\sigma_{A_{est}}^2(\boldsymbol{\theta}) = \frac{1}{2\,n_g^2} \sum_k |\gamma_t(\boldsymbol{\theta}; \boldsymbol{\theta_k})|^2 \, w_k^2 \, \Psi^2(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|). \tag{3.22}$$

This noise estimate is only due to galaxy noise, to get the complete noise (with the LSS component) we must add the constant:

$$C_{LSS} = \frac{1}{2\pi} \int_0^\infty |\tilde{\Psi}(\boldsymbol{k})|^2 P_\gamma(k) k \, \mathrm{d}k \tag{3.23}$$

The sum of equations (3.22) and (3.23) is the *squared noise map*. The *Signal-to-noise ratio map* is defined as the signal map (3.18) divided by the square root of the squared noise map, and it is the seventh and last HDU.

If the calibration factors $m$ are given then the *calibration map* is computed as:

$$c_m(\boldsymbol{\theta}) = \frac{\sum_k (1 + m_k)\, w_k\, \Psi(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|)}{\sum_k w_k \Psi(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|)}. \tag{3.24}$$

The calibration is then included by dividing the equations (3.18) and (3.19) by the calibration map and (3.20) and (3.22) by the square of the calibration map.

### 3.2.3 On the execution of Get the Halo 2021

A correct execution of *Get the Halo* 2021 from the command line requires the name of the setup file that contains all the input parameters. If the program is executed without it, it just prints a set of instructions on how perform a proper execution. If the program is correctly executed then it defines an object of the class *initializer* that reads the names and values of the input parameters from the setup file. Thanks to a method *initializer::get_parameters* those are stored in apposite variables defined in the global namespace so they can be used by the program. A default value is given to the optional parameters that have not been set.

After the initialization section, GtH21 must read the weak lensing data. This is done thanks to two classes, *galaxy* and *catalogue*. The class galaxy stores the weak lensing data of one galaxy, like its position, shear, redshift, weight and bias, and the methods to compute the tangential and radial shear of a given point with respect to the galaxy object. The class catalogue is instead used to hold a collection of galaxy objects, and also to compute some useful quantities that characterise the dataset, such as the number of galaxies, the field of view in the two sky directions, the average number density of galaxies, the average galaxy separation, the field coverage fraction. If the lensing weights are given then GtH21 also computes *effective* variants of the number of galaxies, number density, average galaxy separation, shear modulus mean value, shear variance. If the input parameter x_is_RA is set to 1 the program applies a flat sky approximation to the sky coordinates read from the weak lensing data file. When GtH21 calls the method *catalogue::read_galaxies* the weak lensing data is read and stored into the galaxy objects of the catalogue and its relevant quantities are computed.

The next step in the program execution is the setup of the filter. The filter is read from its ASCII file by declaring and defining a *CosmoBolognalib* object *cbl::data::Table*, that stores in *vector objects* the points of radial distance and filter values where the filter is defined. When the output maps are computed the filter value used are linearly interpolated from the values contained in the filter vector; the interpolation algorithm requires that the radial points where the filter is sampled are all equidistant in order to make the interpolation faster.

The filter is then stored inside apposite members of an instance of the *filtering* class. This class was made to store the filter function and the halo shear profile, it contains the

methods to compute the output maps and to interpolate the filter and the halo signal at any radial position. When the filter is set and stored into members of the filtering object, apposite methods compute the filter typical scale (the weighted average of the radius with the filter) and the cut radius where the filter becomes consistent with 0. This cut radius is always between $60'$ and $120'$.

A crucial step of the program is the setting of the output maps dimensions. If the input parameters *nx* has been set to 0, the method *filtering::optimal_sampling* computes *nx* and *ny* so that the filter typical scale is sampled 5 times (this value could be changed). If *nx* was not 0 it is left unchanged and *ny* is set to respect the field of view proportions. Once the number of pixels of the output maps has been set, they are defined as bi-dimensional pointers of double.

The galaxy number density map is the first to be computed thanks to a method *catalogue::make_mask*; it creates a grid where the number density of galaxies is defined. This map is used as a mask, so that no computational time is spent into computing the output maps where there are no galaxies.

Now that everything has been set, the output maps can be computed thanks to method *filtering::M_ap* (*Aperture mass*). This is the step that takes most of the computational time.

After this step some output properties are computed, such as the mean values of the square root of the variance map, the variance of the E-mode and B-mode maps and the theoretical E-map variance from galaxies noise. Those and others will be saved into the header of the primary HDU.

In the next step of the program the constant defined in equation (3.23) is computed and added to the variance map and the signal to noise ratio map is produced.

The final step of the program is writing the output maps into the FITS file and saving the relevant quantities about input parameters, catalogue properties and map features into the header of the primary HDU, including WCS coordinates.

### 3.2.4   On the halo detection algorithm

Once the program is run, the FITS file described in section 3.2.2 is produced. A first rapid check for detection can be done by looking at the E-map produced in the primary HDU of the output file, or in the S/N map. The biggest haloes can be seen as spots of high values of the E-mode estimator defined in equation (3.18) surrounded by a trough of lower values, as can be seen in figure 3.3. This is the typical signature of the optimal weak lensing filter, that is characterised by having a peak (that produces the peak spot) and a slightly negative region just outside the peak (responsible for the trough). The values of the estimator can either be positive or negative; as a general trend negative regions tends to be more devoid of matter, while positive regions tends to contain more matter. Of course, since the data always contain noise this is a description of a general trend, not a fixed rule. The distribution of the E-mode values and the S/N that the

pixels of an output map present is made up by two structures, as it is shown in figure 3.4. The first is an almost symmetrical distribution centered around 0 (if there are not systematical errors), that is given by the pixels that contain only noise. The second is a high-values tail that is made up by the pixels that contain the detected haloes.
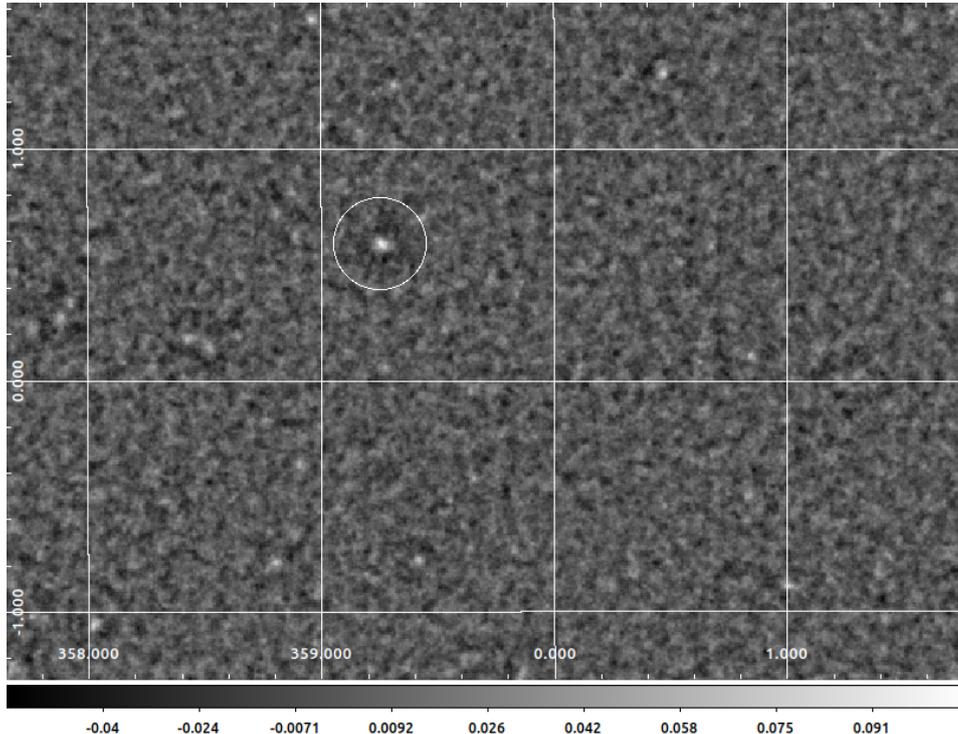


**Figure 3.3:** Part of an output map from a set of simulated data that will be described in chapter 4. The picture shows in a grey scale the E-mode, and a large cluster shown inside a circle can be seen as a white spot surrounded by a darker trough.

The algorithm we used relies on being able to observe the high value tail of the detections and separate it from the noise in order to select the pixels that contain the halo candidates. The pixels that contain noise are identified by evaluating the B-mode on the map thanks to equation (3.19) that uses the cross shear. As described in sections 2.1.4 and 2.2.1, noise has both E and B modes, while true weak lensing signal contains only E-mode. The distribution of the S/N of the B-mode of an output map contains the pixel distribution around 0, but not the high values tail (the signal from the detected haloes); it can thus define the range of signal values that can be produced just by noise. The detection tail is found above those values.

Let us note that there are two possible operative definition of the cross shear. If we rotate each galaxy shear by $\pi/4$ and evaluate on this data an estimator of lensing signal, such as the one defined in equation(3.18), we get a result consistent with 0 since we
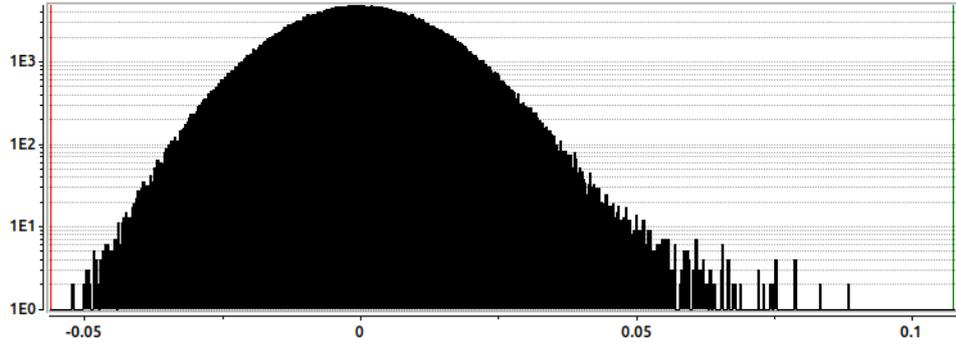
**Figure 3.4:** Distribution of the values of the E-mode of an output map from simulated data that will be described in chapter 4. Most of the pixels have values that belong to a symmetrical distribution around zero: these are pixels corresponding to the noise. A tail at the highest values can be seen, given by the pixels that contain the detections.

have exchanged E and B modes: after the rotation we are using the cross shear and not the tangential shear. The rotation can either be *counter-clockwise* as it was computed in equation (2.68), or *clockwise*, a rotation of $-\pi/4$, i.e. each galaxy is assigned a new complex shear $\gamma'$ with components given by:

$$\gamma'_1 = -\gamma_2, \quad \gamma'_2 = +\gamma_1. \tag{3.25}$$

Both equations (2.68) and (3.25) can be used to define the cross shear. If we compute the aperture mass estimators with shear catalogues rotated once by $+\pi/4$ and once by $-\pi/4$, the two output maps will be perfectly opposites. Since the distribution of the signal we expect from a cross shear is symmetric around zero, they can both be used to estimate the B-mode, consistent with the noise of the E-mode. For instance, the estimator (3.19) is defined considering the cross shear as a $+\pi/4$ rotation according to equation (2.68); instead Umetsu (2020) acknowledges both definitions, but tends to use equation (3.25) when talking about shear rotation and extraction of the B-mode from data.

We decided to be consistent with the convention of Umetsu (2020), hence after GtH21 produced its output, we run it again on the same set of data, with the only difference that the shear components of each galaxy are modified according to equation (3.25). The signal output map we get is of course perfectly opposite to the B-mode we got before, but this new execution of the code is necessary to recompute the noise estimator (3.22) for the cross shear map. The new S/N map we compute can be used as a noise estimator.

We define the pixels that are part of a halo candidate detection as those pixels whose S/N is higher than a given threshold. We used two different thresholds and produced two catalogues for each computed map. For the sake of clarity, we shall call T-map ("true" map) the S/N map obtained using the correct shear, that extracts the E-mode described by the tangential shear. We shall call R-map ("rotated" map) the S/N map obtained

using the rotated shear according to equation (3.25), that extracts the B-mode described by the cross shear. An example of the S/N distributions discussed is given in figure 3.5, while a graphical example of the definition of the two thresholds can be seen in figure 3.6; we defined the two thresholds as follows:

- The **strong threshold** is defined as the maximum of the S/N distribution of the of the R-map. The R-map lacks the high values tail given by the haloes and it is made up only by noise. For this reason we choose to define its maximum S/N as the maximum S/N that can be produced by noise in a given set of data. Any pixel above this threshold in the T-map has a S/N so high that it cannot be given by noise and can be considered as a reliable detection.

- The **weak threshold** is defined to loosen up this definition. Figure 3.6 shows the differences of the T-map minus the R-map normalised over their sum bin by bin. When S/N increases the high values tail of the T-map makes sure that an increasing fraction of the total number of occurrences is given by pixels in the T-map, until it reaches 1 after the strong threshold. To define the weak threshold we consider the bins above 0.9, we group them into contiguous bins and take the first bin of the last group. This means that the weak threshold is the S/N value where the T-map detections become the 90% of the total and their fraction never goes down 0.9 from that point onward.

The pixels above the threshold are considered as the ones where there are the halo candidates. Every area of contiguous pixels whose S/N is above the threshold is considered as a detected candidate halo, or a *detection* for short. If a pixel is a square, contiguous pixels have one side or a vertex in common. The position of the detection is the pixel with the maximum S/N of the halo. This simple algorithm allows us to build a catalogue of positions and S/N of candidate haloes detected through weak lensing.
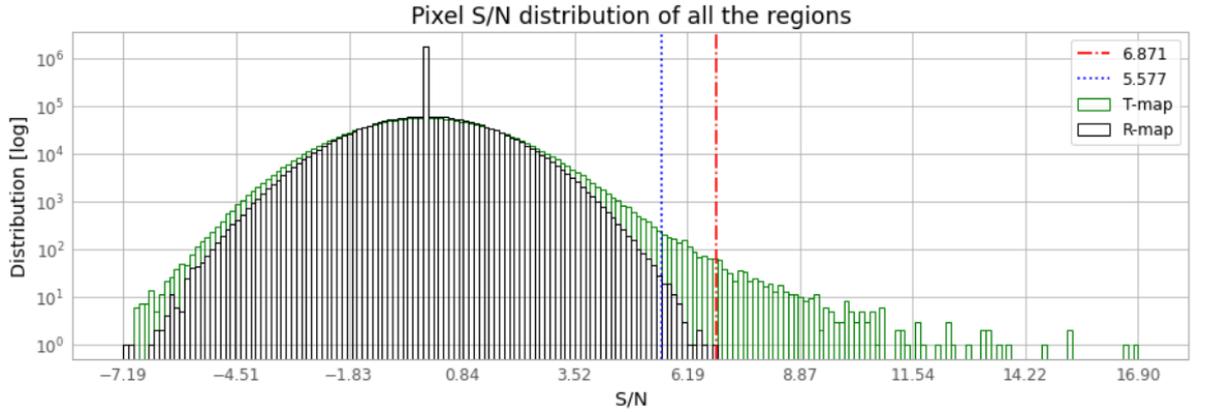
**Figure 3.5:** S/N distribution of the pixels from every region of the output maps of the simulated catalogue discussed in chapter 4. The S/N distribution of the T-map is shown as a green histogram, it is produced by the tangential shear. The S/N distribution of the R-map is shown as a black histogram, it is produced by the cross shear. It is possible to see that the R-map is approximately symmetrical around 0; that is all caused by the B-modes of the noise since there is no lensing signal in the cross shear. The T-map shows a central distribution symmetrical around 0 and with an extension comparable to the distribution of the R-map caused by the E-modes of the noise plus a tail at high values that is not caused by noise but by the E-mode of the lensing signal, i.e. by the detections. The red line is the strong threshold, the blue line is the weak threshold.
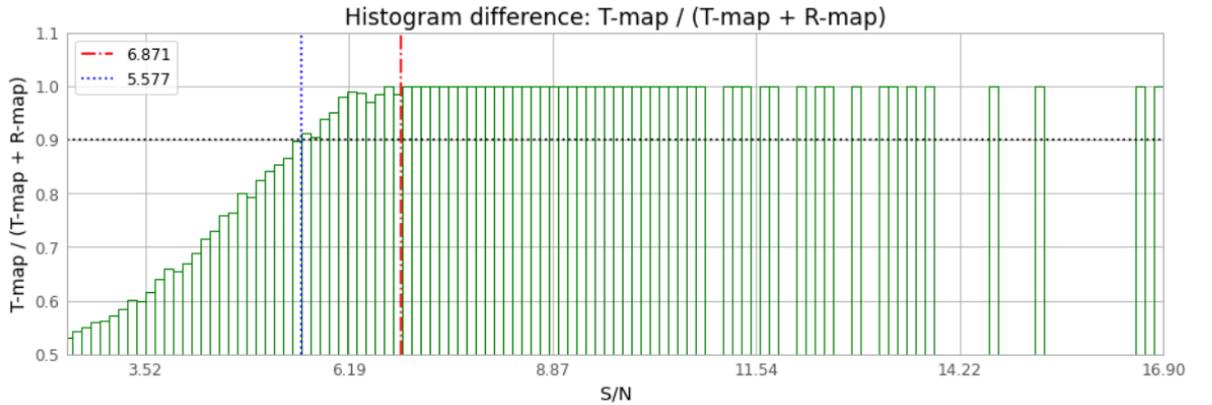


**Figure 3.6:** S/N distribution of the pixels from every region of the output maps of the simulated catalogue discussed in chapter 4, the same data shown in figure 3.5. It is shown the difference of the counts in each bin of the T-map minus the counts in the same bin of the R-map normalised over their sum bin by bin. The strong threshold (red line) is defined as the maximum S/N value of the R-map, over which the S/N values are given by the lensing signal of the haloes. The weak threshold (blue vertical line) is defined as the S/N where the T-map detections become the 90% of the total (black horizontal line).

# Chapter 4

# Forecasting the filter performances with Euclid simulated data

In this chapter we describe the results of the application of the filter described in chapter 3 implemented through GtH21 to a simulated catalogue of ellipticities of galaxies *weak-lensed* by an ensemble of dark matter haloes and the cosmic large-scale structure. They were simulated to represent typical data that will be available thanks to the *Euclid* mission. In section 4.1 we present the simulated catalogue describing how it was made, why it was made and what are its main properties. In section 4.2 we report the results of applying GtH21 on this catalogue when using an optimal filter initialised with a set of "standard" parameters, used for instance in Maturi et al. (2007). We created two catalogues of detections called catalogues *A* and *B*, that we describe in section 4.2.1. We then matched our detections to the haloes of the simulation and we present the results of our matching in section 4.2.2. In section 4.3 we describe the results of another application of GtH21 on this catalogue, but with a different version of the optimal filter, initialised with a set of "customised" parameters tailored on the properties of the simulated catalogue.

## 4.1 On the simulated catalogue for the Euclid test challenge

We applied the program to a catalogue of galaxy ellipticities (shear) derived by a parent cosmological simulation. The analysed dataset is a catalogue of galaxy ellipticities *weak-lensed* by an ensemble of dark matter haloes and the cosmic large-scale structure, that was realised to represent typical data that will be available thanks to the *Euclid* mission. The simulations were produced by Carlo Giocoli, and the noise in the catalogue of ellipticities was added by Sandrine Pires. These simulated data were produced as a part of a "test challenge" organised inside the *Euclid SWG Clusters of Galaxies* and more specifically as

57

a part of the activities of its WP 10, "Weak Lensing Selected Clusters". The goal of the challenge is to select the best cluster detection algorithm that uses weak lensing data, by testing several detection methods on the same dataset. It might be used to detect galaxy clusters as a complementary algorithm to the main ones that are going to be implemented in the Euclid data analysis pipeline. The main algorithms chosen to detect clusters in the Euclid mission use photometric catalogues, so adding an algorithm that uses catalogues of galaxy shear might increase the reliability of the detections if the same ones are found. These algorithms are the *Adaptive Matched Identifier of Clustered Objects* (*AMICO*) code based on matched filtering (Bellagamba et al. 2018) and the *PZWav* code based on an adaptive wavelet approach (Euclid Collaboration et al. 2019).

**The Euclid mission.** Euclid (Scaramella et al. 2021) is a visible to near-infrared space telescope currently under development by the European Space Agency (ESA) and the Euclid Consortium; the launch is scheduled to occur at the end of 2022. See the EUCLID website [1] for more information. The mission will investigate the distance-redshift relationship and the evolution of cosmic structures by measuring shapes and redshifts of galaxies and clusters of galaxies out to redshifts $\sim 2$ (a look-back time of $\approx 10 \, \mathrm{Gyr}$). In this way, Euclid will cover the entire period over which dark energy played a significant role in accelerating the expansion of the Universe: the objective of the Euclid mission is in fact to better understand dark energy and dark matter, that make up most of the energy content of the Universe, as seen in chapter 1. In particular, Euclid is expected to help answering questions such as:

- What is the distribution of dark matter in the Universe?

- What is the history of expansion of the Universe?

- What does this tell us about the nature of dark energy? Does the dark energy equation of state evolve over time?

- How do large-scale structures form in the Universe?

Euclid is optimised for two primary cosmological probes:

- *Weak gravitational Lensing* by measuring distortions of galaxy images by mass inhomogeneities along the line-of-sight, in order to map the dark matter distribution and determine the effect of dark energy.

- *Baryonic Acoustic Oscillations*, wiggle patterns imprinted in the clustering of galaxies on large scales, which provide a standard ruler to measure the expansion of the Universe and the acceleration caused by dark energy.

---

[1] https://sci.esa.int/web/euclid

Euclid will survey the sky in a "step and stare" mode: the telescope will point to a position on the sky, with imaging and spectroscopic measurements performed on an area of $\sim 0.5 \deg^2$ around this position. The sensitivity is 25 mag for the visual imager and 24 mag for the near-infrared photometer. The sky coverage strategy is driven by the wide-survey requirement to cover $15\,000 \deg^2$ of extragalactic sky (more than 35% percent of the celestial sphere) during the mission lifetime of 6 years. Additionally, a deep survey will be performed in three deep fields ($40 \deg^2$ in total), reaching two magnitudes deeper than the wide survey.

**The numerical simulations.**    The parent cosmological simulation used to produce the simulated catalogue is the result of a dark-matter only N-body simulation, carried out with the code *GADGET-2* (Springel 2005; Springel et al. 2001a) by Carlo Giocoli. The simulation assumes a $\Lambda$-CDM model with normalized Hubble constant $h = 0.67$, dark matter density parameter $\Omega_{CDM} = 0.27$, baryon density parameter $\Omega_{bar} = 0.05$, curvature density parameter $\Omega_K = 0$ (flat Universe), normalization of the power spectrum of the initial density fluctuations $A_s = 2.1265 \times 10^{-9}$ (used instead of $\sigma_8$, the rms density fluctuations in spheres of $8 \,\mathrm{Mpc}\, h^{-1}$), spectral index of the power spectrum of the initial density fluctuations $n_s = 0.96$, current temperature of the CMB radiation $T_{CMB} = 2.7225 \,\mathrm{K}$, effective number of relativistic species (degrees of freedom) $N_{eff} = 3.046$. The simulated box is a cube with a side length of $1 \,\mathrm{Gpc}\, h^{-1}$. It contains $1024^3$ dark matter particles that evolve from redshift 99 to 0. Haloes have been identified at each redshift by using *SUBFIND* (Springel et al. 2001b). Aiming at studying light propagation through an inhomogeneous universe, the authors have taken several snapshots of the simulation at different redshifts; those snapshots have been stacked in order to construct light-cones for the following ray-tracing analysis. The light cones of the simulations have been built with *MapSim* (Giocoli et al. 2018, 2015); this method has been validated by comparison with other programs (Hilbert et al. 2020). *MapSim* requires to set the desired field of view, chosen to be a light cone with a 10 deg aperture, and the number of snapshots ahead of time; the authors saved 43 snapshots from redshift 0 to 4. Redshift 4 was chosen to better understand where the dynamical evolution of the matter-energy components of the Universe starts to leave a mark in the weak lensing observables (Giocoli et al. 2015). Each snapshot consists of a cubic volume containing one realization of the matter distribution at a given redshift. However, since they are all obtained from the same initial conditions, these volumes contain the same cosmic structures at different stages of their evolution, approximately at the same positions in each box. In order to avoid repetitions of the same cosmic structures along one line-of-sight, the snapshots cannot simply be stacked as a consecutive sequence, they need to be randomised. The randomization process is described in Roncarelli et al. (2007). The snapshots are converted from being three-dimensional volumes (containing one realization of the matter distribution at a given redshift) into a two-dimensional mass distribution, by projecting the particle positions

to the nearest pre-determined plane, maintaining their angular positions. Such planes were used as 43 lens planes for the ray-tracing simulations. Those simulations used the ray-tracing code *GLAMER* (Metcalf et al. 2014; Petkova et al. 2014). The light cone has been populated with sources at random right ascension and declination, but with a redshift distribution in the range $[0, 3]$ and peaked at redshift $\approx 1$, a typical distribution that we expect to see with Euclid. The ray-tracing simulations produced as output maps of convergence and shear. We use the shear catalogues produced by these simulations.

**The simulated data.**  The simulated data we used consist of a catalogue of $10\,801\,552$ galaxy positions and ellipticities, randomly distributed in one field of $10\,\mathrm{deg} \times 10\,\mathrm{deg}$, with right ascension and declination both in the range $[-55]\,\mathrm{deg}$. Their main properties are shown in figure 4.1. The distributions of the two ellipticities components are Gaussian with 0 mean and 0.26 standard deviation; they are shown in figure 4.1a. The redshift distribution is shown in figure 4.1b; the median is 0.826, the quartiles are $[1.207 \div 0.523]$, the redshift range of the galaxies is $[0, 3]$. This catalogue simulates typical data that we expect to be measured in a survey of Euclid mission. Let us note that there is a slight difference: the wide survey of the Euclid mission is expected to detect galaxies up until redshift $\sim 2$; the simulated catalogue up until redshift 3.

The galaxies show ellipticity not only because of their own intrinsic ellipticity, but also because of the optical deformation due to weak lensing by haloes of clusters of galaxies simulated in the field. There are 1611 haloes in the field, their properties are shown in figure 4.2. Their redshift distribution, shown in figure 4.2a, goes up to redshift 1. The virial mass distribution of the haloes is shown in figure 4.2b. All the masses considered in this chapter are $M_{200}$. The haloes have masses in the range $[7.4 \times 10^{13}, 1.4 \times 10^{15}]\,\mathrm{M}_\odot\,h^{-1}$, with median $1.07 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$ and quartiles at $[0.86 \div 1.47] \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$. There are only 31 haloes (2% of the total) with mass $> 4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$ and only 8 with mass $> 6 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$. Most of the haloes have then intermediate mass; for instance, 1190 haloes (74% of the total) have mass in the range $[0.75, 1.5] \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$.

When GtH21 is applied on the shear catalogue it gives some extra information on the data (before even reading the filter), such as:

- Galaxy number surface density: $30.0\,\mathrm{arcmin}^{-2}$;

- Average galaxy separation on the sky: $0.182'$;

- Mean shear modulus: 0.33;

- Root mean square of the shear modulus: 0.36

- White noise from galaxies: $P_\epsilon = \frac{1}{2}\frac{\sigma_{\epsilon_s}^2}{n_g} = 1.88 \times 10^{-10}\,\mathrm{rad}^2$.

These quantities can be taken into account when setting the parameters to build the optimal filters, as discussed in section 3.2.1.

**(a)** Ellipticity distributions.



**(b)** Sources redshift distribution.
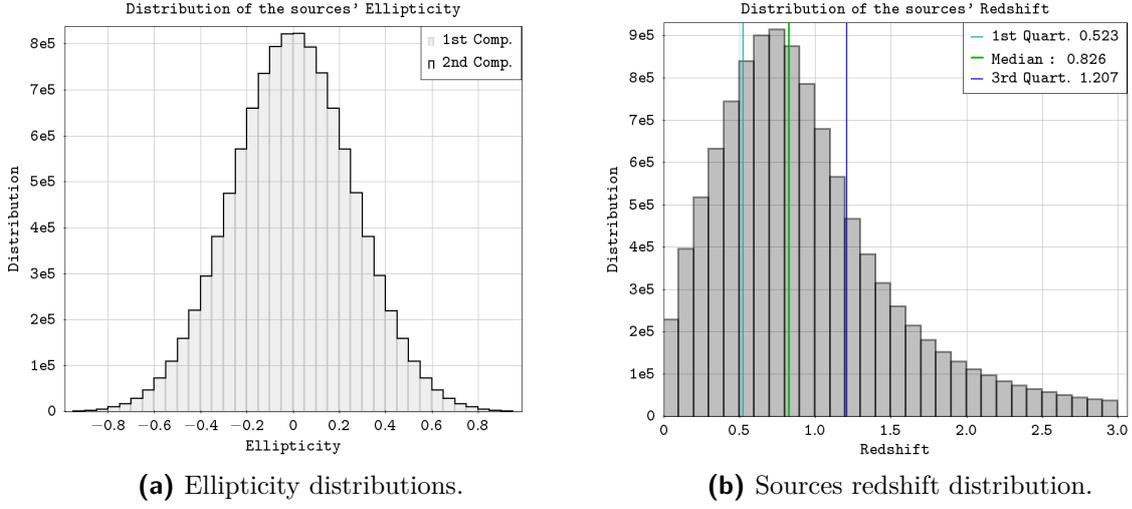
**Figure 4.1:** Distributions of ellipticity and redshift of the simulated sources in the "Euclid test challenge" shear catalogue. In **figure 4.1a** we show the distribution of the first (grey semi-filled histogram) and second (black step histogram) ellipticity components. They are both gaussian with average 0 and standard deviation 0.26, the two distribution are consistent with each other. In **figure 4.1b** the redshift distribution of the simulated galaxies is shown. The first quartile, median and third quartile of the distribution (0.523, 0.826, 1.207) are shown by the vertical coloured lines.



**(a)** Haloes redshift distribution.



**(b)** Virial mass distribution.
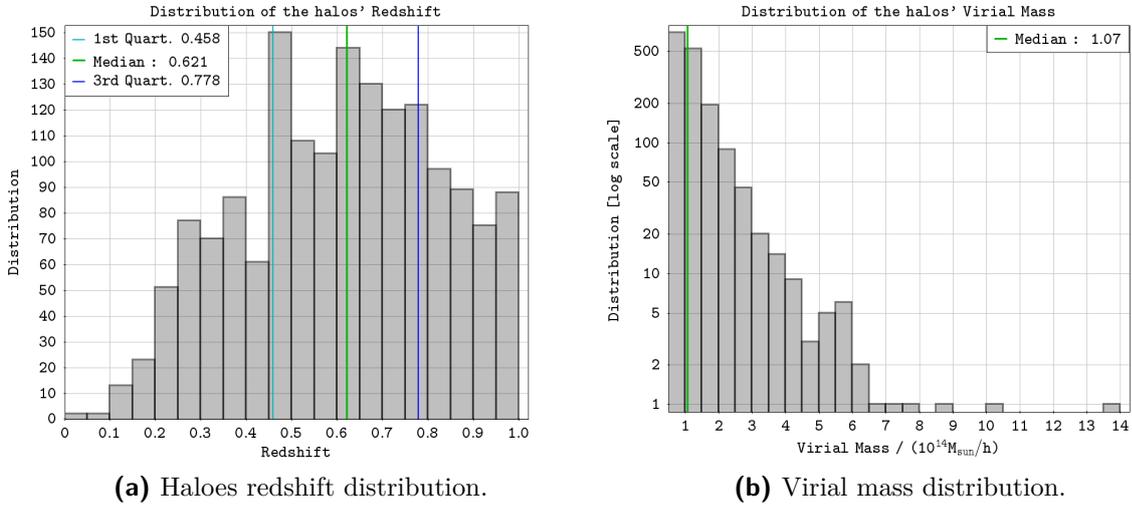
**Figure 4.2:** Distribution of redshift and virial mass ($M_{200}$) of the 1611 simulated haloes in the "Euclid test challenge" halo catalogue. The redshift distribution is shown in **figure 4.2a**, with the first quartile, median and third quartile at 0.458, 0.621, 0.778 respectively as coloured lines. The virial mass distribution is shown in **figure 4.2b**, with the median at $1.1 \times 10^{14} \, \mathrm{M_\odot} \, h^{-1}$ (green line).

## 4.2 Application of a "standard" filter on the catalogue

We analysed the data by using an optimal filter with the following parameters:

- Lens mass: $1.0 \times 10^{15} \, \mathrm{M}_\odot \, h^{-1}$

- Lens redshift: 0.3

- Source redshift: 1.0.

- White galaxy noise: $1.28 \times 10^{-10} \, \mathrm{rad}^2$.

- Beam FWHM: $0.3'$.

This is a set of "standard" parameters, used for instance in Maturi et al. (2007), meant to detect high-mass haloes that have intermediate distance between observer and sources, a condition that increases lensing effects. This is due to the dependence of the geometrical lensing strength on the angular-diameter distances between the observer and the lens, the lens and the sources, and the observer and the sources (Pace et al. 2007). This filter has cut radius $46.68'$, that GtH21 extends to $60'$: only galaxies closer than $60'$ from a given pixel in the output maps are taken into account to compute the values of the output statistics, as described in section 3.2.2. The average radius of this filter is $2.2'$, so each pixel is large $1/5$ of this scale, i.e. $0.44'$. Some information on the numerical application on the data is given in figure 4.3. In particular, the filter used is shown in figures 4.3a (real space) and 4.3b (Fourier space). In order to reduce the execution time of GtH21 on the computer used for the analysis we divided the shear catalogue into nine sub-catalogues, each one representing a different spatial region. The division into sub-fields is shown in figures 4.3c and 4.3d. When we run the program on each sub-catalogue we included galaxies from an extra amount of $60'$ both in the right ascension and declination to make sure that the division would not introduce border effects in the maps. For instance, the central region of the field had RA $\in [-1.5; 1.5] \, \mathrm{deg}$, DEC $\in [-1.5; 1.5] \, \mathrm{deg}$. In order to correctly compute all the estimators in every pixel of the map, even the ones near the borders, we included in the input file all the galaxies with RA $\in [-2.5; 2.5] \, \mathrm{deg}$, DEC $\in [-2.5; 2.5] \, \mathrm{deg}$. In this way we are certain that if we processed the entire catalogue in one run we would get the same output maps.

After running the program we obtained nine output files. The E-mode and B-mode of the output map of the central region is shown in figure 4.4 as an example of a typical output. We then computed the strong and weak S/N thresholds above which a pixel is assumed to contain a detection as described in section 3.2.4; this is shown in figure 4.5. We computed the histogram of the S/N distribution of the pixels of each sub-field with the same binning and we summed them to get the S/N distribution of the S/N of all the

pixels of the entire field of the simulation; we did this for the T-maps and the R-maps. The sums are then the distributions of the pixel S/N for all the regions of the simulation, that we plot in figure 4.5a for both the T-map and the R-map. Let us remember that we defined the strong threshold as the maximum S/N of the R-map; for our data and the filter described in this section it is 6.871, as it is shown in panel 4.5a. In figure 4.5b we use the same binning to plot, for each bin, the ratio between the counts of the T-map divided by the total number of counts (T-map plus R-map). This is used to identify the weak threshold, since we defined it as the S/N value above which the T-map detections are always more than 90% of the total (T-map plus R-map). For our data and the filter considered in this section it is 5.577 as it is shown in figure 4.5b.

Once the S/N thresholds were defined, we run the halo detection algorithm as described in section 3.2.4. We defined the pixels that are part of a *detected halo-candidate* (*detection* for short) as those pixels whose S/N is higher than a given threshold. Every group of contiguous pixels (they must have a vertex or a side in common) whose S/N is above the threshold is considered as a detection. The centre of the detection is assumed to be the pixel with the maximum S/N of the group; that S/N is the S/N of the detection. With this algorithm we have built two catalogues of detections; we included in a catalogue for each detection its position, its S/N, the number of pixels covered by it, its area on the sky in arcmin (computed as the number of pixels times the area of one pixel in arcmin).

Since we have two thresholds, we have built two catalogues. We shall refer to the catalogue built with the strong threshold as **catalogue A** and the one built with the weak threshold as **catalogue B**.

**(a)** "Standard" optimal filter in real domain.

**(b)** "Standard" optimal filter in the Fourier space.

**(c)** Field division: 4 sub-fields shown.

**(d)** Field division: 5 sub-fields shown.
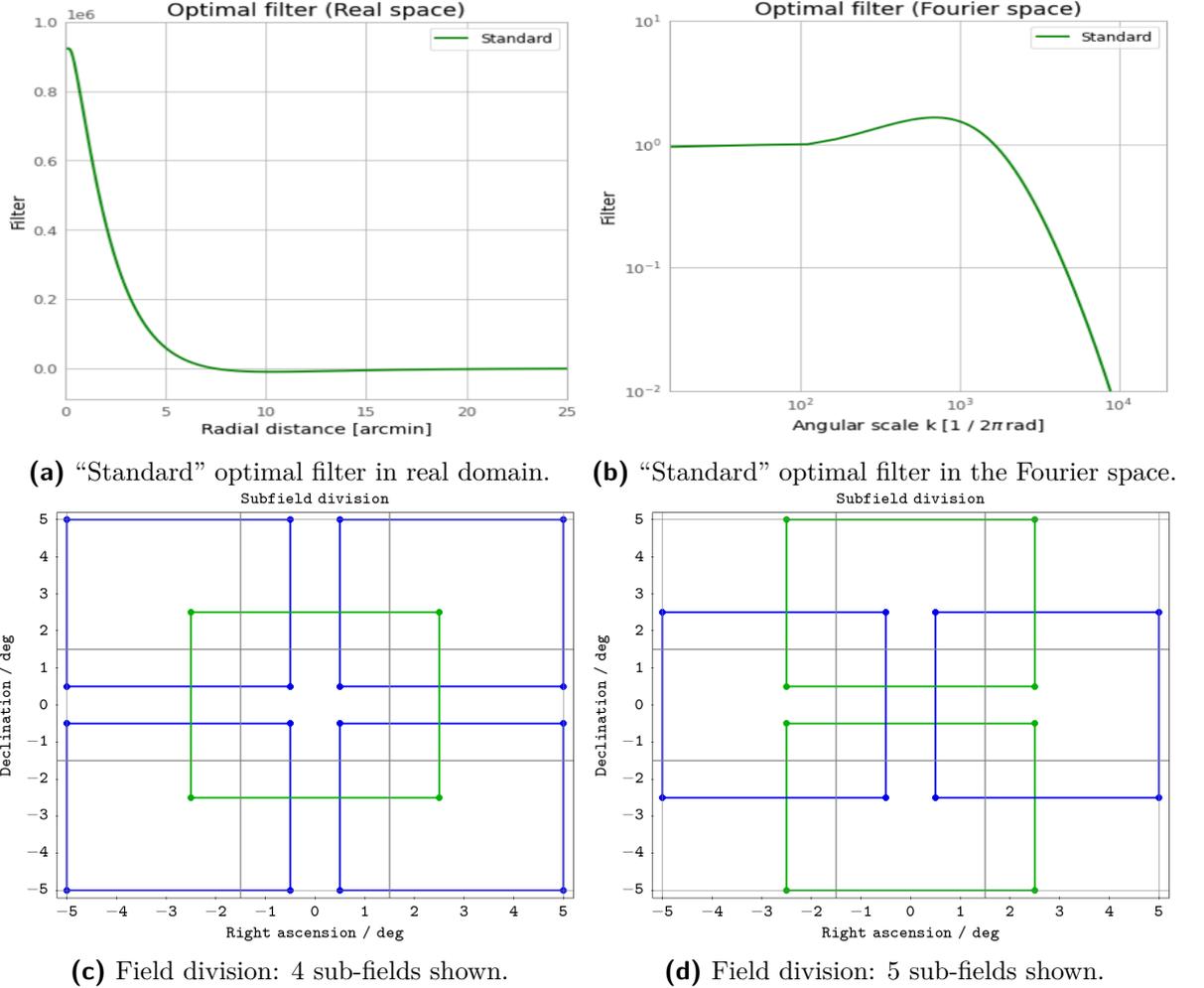
**Figure 4.3:** Application of GtH21 on the data. The "standard" optimal filter applied on the data and discussed in section 4.2 is shown in the real space in **panel 4.3a**, and in the Fourier space in **panel 4.3b**. The division into nine sub-fields necessary to reduce the computational time on the machine we used is shown in **panels 4.3c** and **4.3d**. The three ranges used are $[-5, -1.5]$ deg, $[-1.5, 1.5]$ deg, $[1.5, 5]$ deg as shown by the grey lines in both panels. They divide the field into nine sub-fields with no superposition. When we run GtH21 on each sub-catalogue we included galaxies from an extra amount of $1$ deg to make sure that the division would not introduce border effects in the maps. When we add the extra degree we obtain nine regions with partial superposition, described by the nine coloured rectangles, four of which are shown in **figure 4.3c** and five in **figure 4.3d**.
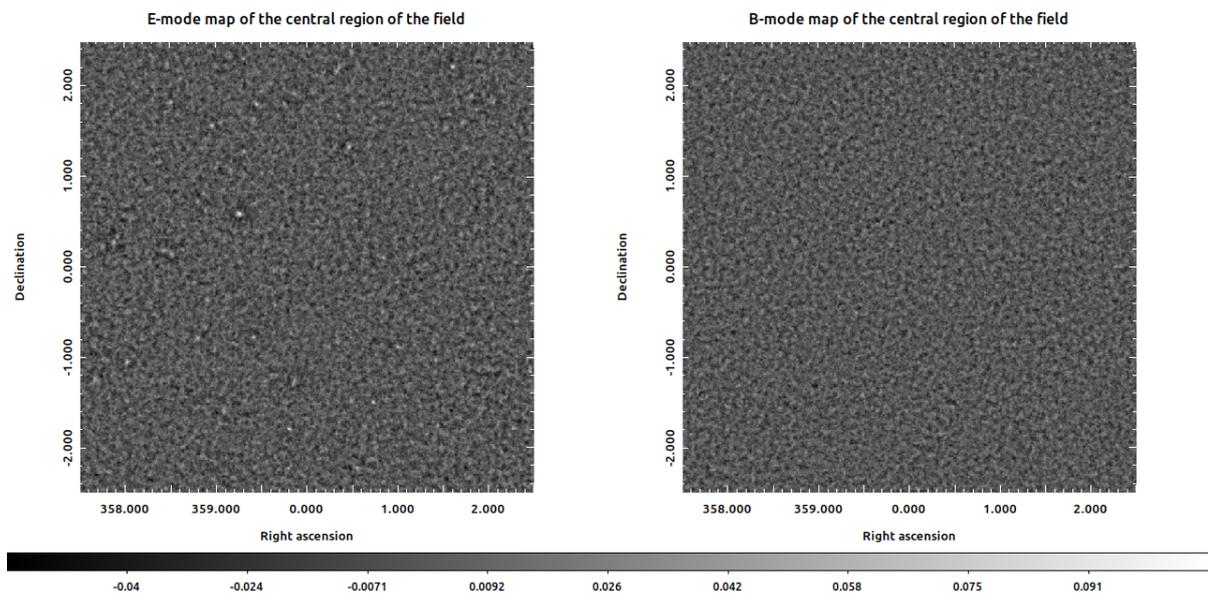
**Figure 4.4:** Output maps of the central region of the simulation, the E-mode is on the left, the B-mode on the right. They are plotted with the same colour scale. In the E-mode map it is possible to see the typical signatures of the presence of haloes: high-value spots surrounded by lower-value troughs. These are not present in the B-mode map, that only contains the noise from galaxies and the large scale structure.

**(a)** S/N Distribution



**(b)** Threshold identifier
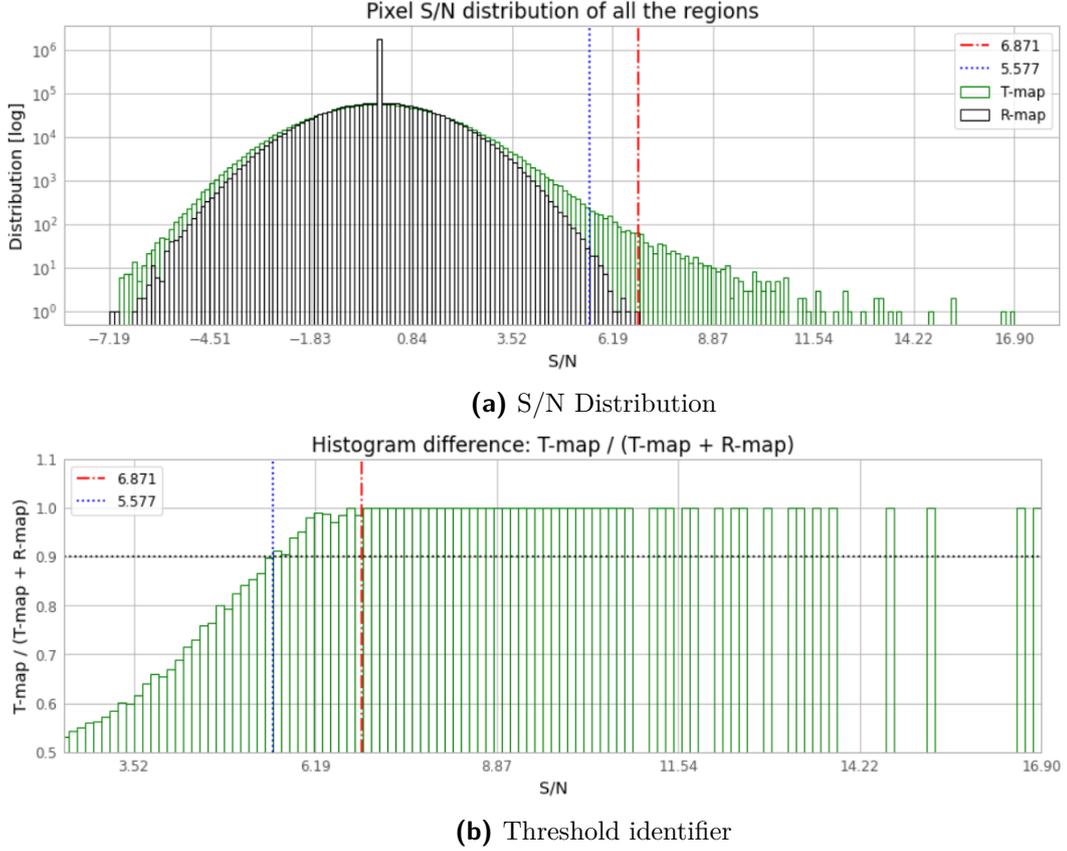
**Figure 4.5:** Histograms used to define the strong and weak thresholds, as described in section 3.2.4. **Panel 4.5a** shows the distribution of the S/N of all the pixels of the entire field of the simulation. The green histogram is the T-map (S/N of the E-mode, that uses the tangential shear), the black histogram is the R-map (S/N of the B-mode, that uses the cross shear). It is possible to see the high-S/N tail of the detections. The strong threshold is the maximum S/N of the R-map. For our data and the filter considered in section 4.2 it is 6.871; it is shown as a red line on both panels. In **panel 4.5b** we use the same binning as the upper panel to plot, for each bin, the ratio between the counts of the T-map divided by the total number of counts (T-map plus R-map). This is used to identify the weak threshold, since we defined it as the S/N value above which the T-map detections are always more than 90% of the total (T-map plus R-map). For our data and the filter considered in section 4.2 it is 5.577; it is shown as a blue line on both panels.

## 4.2.1 Catalogues A and B

Catalogue $A$ is built with a S/N threshold of 6.871, it is made up by 122 detections. Catalogue $B$ is built with a S/N threshold of 5.577, it is made up by 449 detections. Catalogue $B$ can be thought as an extension of $A$ at lower S/N, so we expect every detection in $A$ to also be found $B$, with the same position and S/N. The ID numbers of the same detection in the two catalogues will in general be different, and the area of a detection in cat. $B$ will be larger than the area of the same detection in $A$, because we are allowing pixels with S/N between the two thresholds to be considered as part of a detection in $B$. It is *mostly* true that a detection of cat. $A$ is also found in $B$, but there are two exceptions: detections $A\#87$ and $A\#97$, i.e. the detections of cat. $A$ with ID equal to 87 and 97 respectively, are not in cat. $B$. This is because they are very close to stronger S/N peaks, sources $A\#88$ and $A\#96$. These two sources are found in cat. $B$ as $B\#341$ and $B\#355$ respectively. Their area is so large in cat. $B$ that they include the pixels that produce, respectively, $A\#87$ and $A\#97$. Hence the pixels that form detections $A\#87$ and $A\#88$ are both included in the set of pixels that make up $B\#341$, but the position of the detection is the one of the stronger peak, i.e. $B\#88$. The same goes for $A\#96$, $A\#97$ and $B\#355$. We believe that detections $A\#87$ and $A\#88$ (same for $A\#96$ and $A\#97$) are probably one true detection that has been "fragmented" by the noise and the more restrictive threshold. Our algorithm for halo detections counts only connected pixels as a detection so it considers them to be separate, but a more sophisticated algorithm, for instance a Friends-of-Friends, might count them as one single source (Pace et al. 2007). We report these detections in table 4.1.

The S/N distribution of all the detections of catalogues $A$ and $B$ can be seen in figure 4.6, where it becomes evident that catalogue $B$ is an extension of $A$. The spatial distribution of the detections of the two catalogues can be seen in figure 4.7, where we also plot which detections have been matched to a simulated halo, as it is going to be described in the following section 4.2.2.

**Table 4.1:** Detection $A\#87$, $A\#88$, $A\#96$, $A\#97$ and $B\#341$, $B\#355$ discussed in section 4.2.1. We report the ID of the detections in their catalogues, their position, signal to noise ratio, the area covered by each detection in pixels, their ID in the other catalogue (if they are found in both of them). We also report if the detections have been matched with a simulated halo, as discussed in the following section 4.2.2. The ID number of the halo in the simulated catalogue is given as $H\#$ followed by the ID number.

**(a)** Catalogue $B$.

| ID | RA [deg] | DEC [deg] | S/N | Area [pxl] | ID[cat.A] | Matched Halo |
|----|----------|-----------|-----|------------|-----------|--------------|
| $B\#341$ | 0.09262 | 2.89650 | 7.61784 | 19 | $A\#88$ | - |
| $B\#355$ | -0.09159 | 3.92323 | 6.97951 | 7 | $A\#96$ | $H\#194$ |

**(b)** Catalogue $A$.

| ID | RA [deg] | DEC [deg] | S/N | Area [pxl] | ID[cat.B] | Matched Halo |
|----|----------|-----------|-----|------------|-----------|--------------|
| $A\#87$ | 0.09264 | 2.85978 | 6.94219 | 1 | - | $H\#941$ |
| $A\#88$ | 0.09262 | 2.89650 | 7.61784 | 3 | $B\#341$ | - |
| $A\#96$ | -0.09159 | 3.92323 | 6.97951 | 1 | $B\#355$ | - |
| $A\#97$ | -0.09160 | 3.93789 | 6.97606 | 1 | - | $H\#194$ |
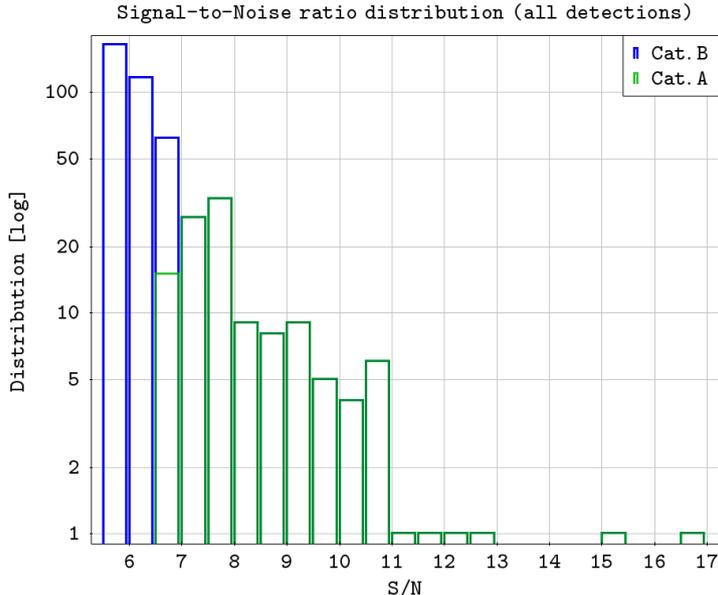
**Figure 4.6:** Signal-to-Noise ratio distribution of all the detections in catalogues *B* (blue histogram) and *A* (green histogram). Catalogue *A* has been superimposed onto *B*; this shows graphically that from the S/N bin $[7.0; 7.5]$ the two catalogues have the same detections, as we expected from our discussion in section 4.2.1.

## 4.2.2 Matching catalogues A and B

For the sake of clarity, we shall refer to the dark matter haloes simulated and described in section 4.1 as *Haloes*, and our candidate-halo detections in catalogues *A* and *B* as *Detections*.

We matched our detections with the simulated haloes by using *TOPCAT* Sky algorithm (Taylor 2005, 2006). *TOPCAT* matches detections of different catalogues by comparing their position on the sky only, given a maximum distance allowed between the two. This is the only parameter needed to match two different catalogues. In general we expect the positions of haloes and their matched detections to be *slightly* different, because GtH21 produces maps defined on a grid of points, that is different from the grid used in the cosmological simulation to determine the position of the haloes. Thus there is a *mismatch* between detections and haloes; we expect the distribution of their distance to be approximately Gaussian. If we set a maximum distance too small then the distribution of the distance between the haloes and their matched detections will appear abruptly cut; if it is too large the distribution will have a peak (approximately Gaussian) plus a small tail at high-distance values of spurious matches. The ideal value for the maximum distance parameter is the one that allows us to see the whole peak in the distribution without the tail of spurious matches.

We tried several values for the *maximum mismatch distance* of the haloes and our

catalogues of detections on the scale of the arcmin (Pace et al. 2007), and we have chosen the value $1.4'$. When using this maximum mismatch distance *TOPCAT* managed to match 100 out of 122 detections (82.0%) of catalogue *A*, while it matches 218 out of 449 detections (48.6%) of catalogue *B*. When a detection is matched to a halo it is given the mass and redshift of the halo, while to the matched halo is given the S/N of the detection. The distribution of the detected and matched sources can be seen in figure 4.7. Figure 4.8 shows the *detection-halo mismatch*, i.e. the distance vector between the matched detections and their halo. Figure 4.8a shows the Right ascension and Declination components of the mismatch vectors, while figure 4.8b shows the modulus of the vector, i.e. the distribution of the halo-detection distance. Figure 4.8a tells us that GtH21, the detection and matching algorithms do not introduce systematic errors in the position of the detections. In fact, the matched detections of cat. *A* have average right ascension mismatch of $0.19' \pm 0.34'$ (error is the standard deviation) and average declination mismatch of $-0.10' \pm 0.59'$. The matched detections of cat. *B* have average right ascension mismatch of $0.21' \pm 0.41'$ and average declination mismatch of $-0.13' \pm 0.61'$. These average shifts are all consistent with zero, so they can be neglected. Figure 4.8b show that the distributions of the mismatch distance are peaked and without tails, so $1.4'$ is a good choice for the maximum mismatch distance allowed. The average value ($\pm$ the standard deviation) of the mismatch distance distribution is $0.64' \pm 0.31'$ for catalogue *A* and $0.70' \pm 0.33'$ for catalogue *B*. The distribution we see in figure 4.8b for cat. *B* shows that the bin $[0.6; 0.8]$ has less counts than $[0.8; 1.0]$; we expected them to be more. This might be a fluctuation due to the poor statistics of the bins.

All the detections matched in catalogue *A* have been matched in catalogue *B* with the same halo of the simulation, except for one detection: source $A\#87$. We discussed in section 4.2.1 that detections $A\#87$, $A\#88$ of cat. *A* are representative of one source (detection $B\#341$ of cat. *B*) that has been fragmented by noise, a more restrictive threshold and our simple halo detection algorithm. These detections were described in table 4.1. *TOPCAT*'s matching algorithm does not take into account the S/N of a detection in the matching process, but only the position of the haloes and the detections. Halo $H\#941$, i.e. the halo with ID number 941 in the simulated halo catalogue, is closer to detection $A\#87$ (distance $0.74$ arcmin) than $A\#88$ (distance $2.93$ arcmin) of cat.*A*, so it gets matched to $A\#87$. In catalogue *B* detection $B\#341$ has the same position of $A\#88$, so it can't be matched to halo $H\#941$ because their distance is larger than the maximum mismatch allowed. The other set of "fragmented" sources discussed in tables 4.1 were $A\#96$ and $A\#97$, $B\#355$ (same position as $A\#96$). Halo $H\#194$ is closer to detection $A\#97$ (distance $0.36$ arcmin) than $A\#96$ (distance $0.57$ arcmin), so it gets matched to $A\#97$. In catalogue *B* detection $B\#355$ has the same position of $A\#96$, so it is matched to halo $H\#194$ because their distance is smaller than the maximum mismatch allowed. So, halo $H\#194$ is matched to detections with different positions in cat.*A* and *B*, while halo $H\#194$ is matched only in cat. *A*. A more sophisticated matching algorithm

that takes S/N and source extension into account would probably match them correctly. Since these are the only "fragmented" sources and they only arise because of the simple detection and matching algorithms we shall not consider them to be statistically relevant for the rest of the chapter: all the other haloes matched with detections in $A$ have been matched with the detection in the same position in $B$, so we can definitively consider cat. $B$ as an extension of $A$ at lower S/N.
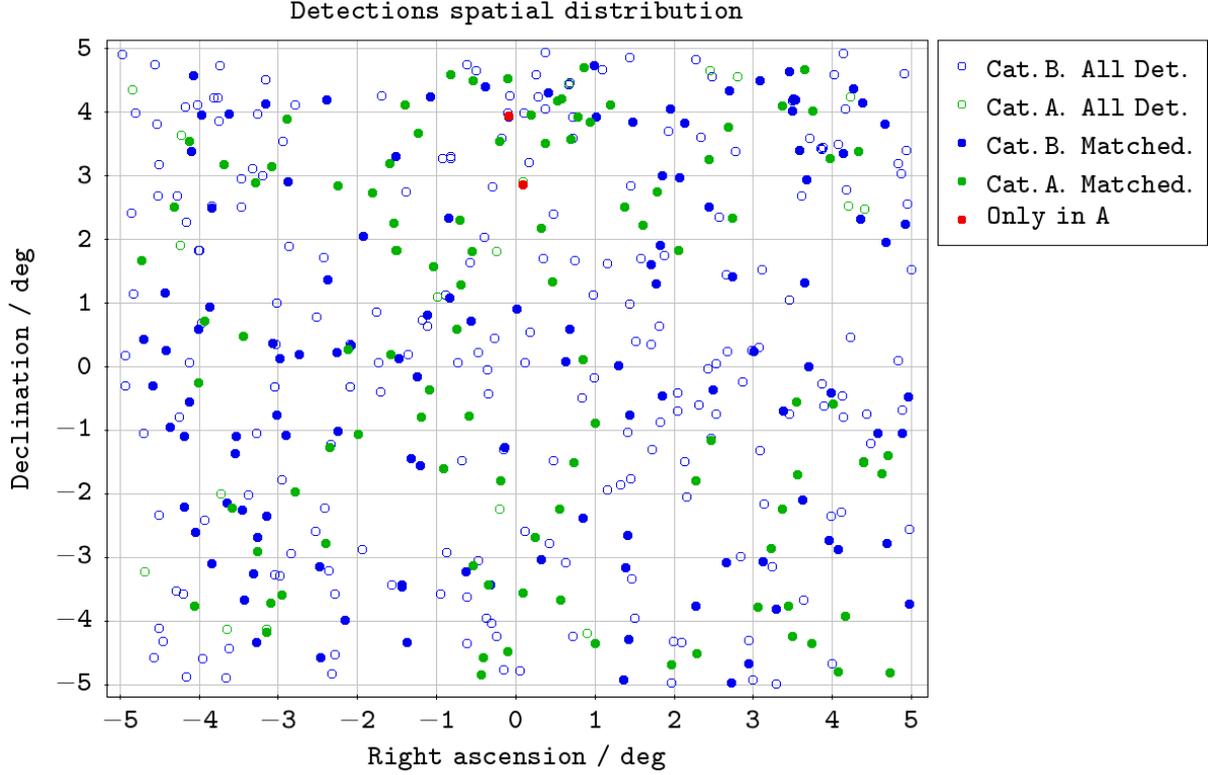


**Figure 4.7:** Spatial distribution of the detections. Green circles are the detections of catalogue $A$ (S/N threshold 6.871). When the circle is empty the detection has not been matched, when the circle is full it has been matched with a halo. These sources are also included in catalogue $B$ (S/N threshold 5.577), with the exception of the two red sources (#87 and #97) discussed in section 4.2.1. Blue circles are the detections of catalogue $B$ only, not found in catalogue $A$. When the circle is empty the detection has not been matched, when the circle is full it has been matched with a halo. We matched our detections with the haloes of the simulation, that have mass in the range $[0.74, 14.0] \times 10^{14} \, M_\odot \, h^{-1}$ and redshift in the range $[0, 1]$.

71

**(a)** Spatial distribution.



**(b)** Distance distribution.
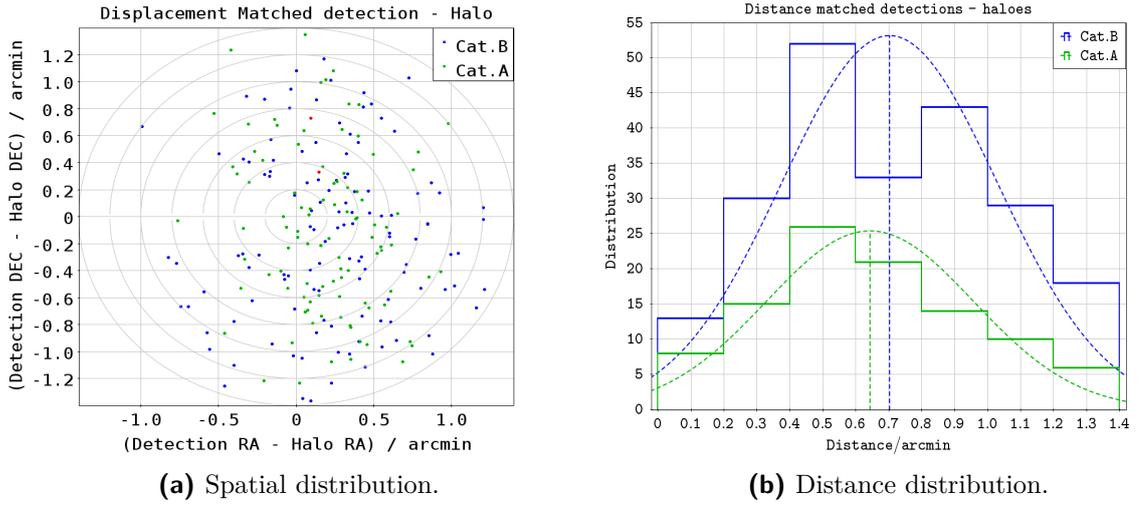
**Figure 4.8:** Distribution of the detection-halo mismatch. We matched our detections with the haloes of the simulation, that have mass in the range $[0.74, 14.0] \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$ and redshift in the range $[0, 1]$. In **panel 4.8a** we show the distribution of the right ascension and declination components of the mismatch vector, defined as the position of the detection minus the position of its matched halo. Green circles are the detections found in both catalogues $A$ and $B$, blue circles only in $B$, red only in $A$. We also plot circles of fixed radius around $(0, 0)$ increasing the radius by 0.2: the counts in each annulus are the heights of the bins shown in **panel 4.8b**. There we show the distribution of the *modulus* of the mismatch, i.e. the detection-halo distance. The green step-histogram is relative to cat. $A$, the blue step-histogram is for cat. $B$. Their maximum value is $1.4'$, the limit set for the matching algorithm. The average and standard deviation of the distributions are $0.64' \pm 0.31'$ for cat. $A$ and $0.70' \pm 0.33'$ for cat. $B$; we show the Gaussian fit of each histogram.

**Signal-to-noise ratio**

The distribution of the S/N can be seen in figure 4.9 for both catalogues *A* and *B*. They show that most of the unmatched detections are introduced when we lower the detection threshold from 6.871 to 5.577. To be more precise, 210 out of the 231 unmatched detections in catalogue *B* have S/N $<$ 6.871. Thus S/N$\simeq$ 7 can be considered as a reliable threshold to detect haloes through weak lensing: 88 out of the 107 detections of catalogue *B* with S/N $>$ 7 are matched. This is consistent with the results of Pace et al. (2007) who found that detections with S/N values greater than $\approx 5 \div 8$ obtained with the optimal linear matched filter of Maturi et al. (2005) can be considered as reliable detections.

We consider an unmatched detection as a *spurious* detection, while we consider a matched detection as a "true" or "pure" detection. This is not the most rigorous definition, since in general spurious detections are peaks caused by noise and true detections are peaks caused by real haloes of the simulations. Since the matching algorithm we used only compares the position of haloes and detections, it is *not impossible* that a random peak caused by a noise fluctuation happens to arise near a halo and is matched to it. We do not expect this to be a relevant problem because several works (Maturi et al. 2005, 2007; Pace et al. 2007) have extensively discussed this topic and shown that the optimal filter discussed in section 3.1.3 is built to maximise the S/N ratio and minimise the spurious detections caused by the large-scale structure, which is the main source of spurious detections. A thorough method to investigate the purity of a catalogue is described by Pace et al. (2007): when a detection is found and matched with a halo of a simulation the authors removed that halo from their catalogue, produced new shear map without the halo and run their program once again. If the detection disappeared it was really caused by the halo (the detection was true), while if it remained then it was caused by the noise (the detection was certainly spurious). We had no access to the simulations, so we could not change them and apply this method.

The *purity* of catalogues *A* and *B*, i.e. the percentage of matched detections in our catalogues, is shown in figure 4.10 where we plot the fraction of matched detections with S/N above a given value. The S/N values at which we compute the points are the lowest end of the bins used in figure 4.9. Above S/N$\simeq$ 7.0 the percentage of matched detections is always above 83%, so this can be considered as a reliable S/N value for the construction of a catalogue. This is consistent with the results of Pace et al. (2007) who found that detections with S/N values greater than $\approx 5 \div 8$ obtained with the optimal linear matched filter of Maturi et al. (2005) can be considered as reliable detections.

The unmatched sources with high S/N are probably spurious detections caused by the large scale structure. Although the optimal filter can minimise the contamination from the large scale structure, as discussed throughout chapter 3, it is not possible to completely eliminate the contamination from the large scale structure since the boundary between it and the virialized haloes is not sharp. Roughly speaking, $\approx 17\%$ of the detected haloes with S/N $>$ 7 might be spurious peaks caused by weak lensing from the
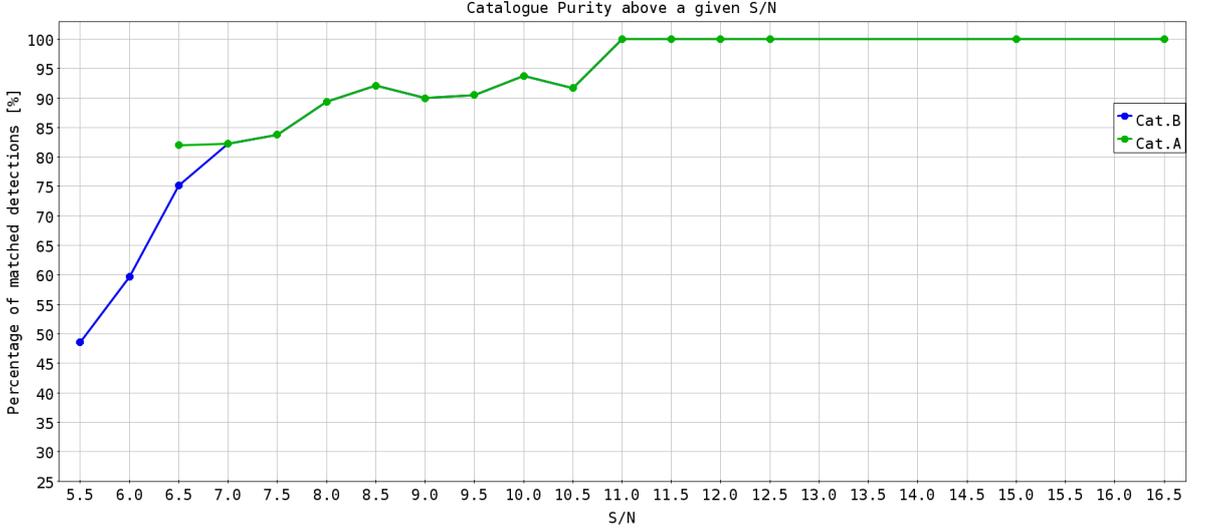
large scale structure.

In figure 4.11 we show the S/N distribution of the matched detections as a function of mass and redshift for catalogues $A$ (figure 4.11a) and $B$ (figure 4.11b). We have chosen a not uniform binning in order to visualise the same mass and subsets that we are going to use later in this section in figures 4.14a, 4.14b and 4.19. The bins are separated at masses $[1, 2, 4] \times 10^{14} \, \mathrm{M_{\odot}} \, h^{-1}$ and redshifts $[0.2, 0.4, 0.6]$. Figures 4.11a and 4.11b show that we find detections with higher S/N in bins of redshift $\in [0.2; 0.4]$ and mass $> 4 \times 10^{14} \, \mathrm{M_{\odot}} \, h^{-1}$. This suggests that our linear filter is optimised to detect haloes with these redshifts and masses. It selects preferentially the haloes around the lens redshift used to build the filter (0.3 in this case) and those haloes with redshift between 0.2 and $0.4 \div 0.5$, that have an intermediate distance between observer and nearby background sources, condition that maximises the lensing effects (Pace et al. 2007). This reflects the dependence of the geometrical lensing strength on the angular-diameter distances between the observer and the lens, the lens and the sources, and the observer and the sources. This result is consistent with what we are going to find in the rest of the chapter.

**(a)** Catalogue *A*.



**(b)** Catalogue *B*.



**(c)** Catalogues *A* and *B*.
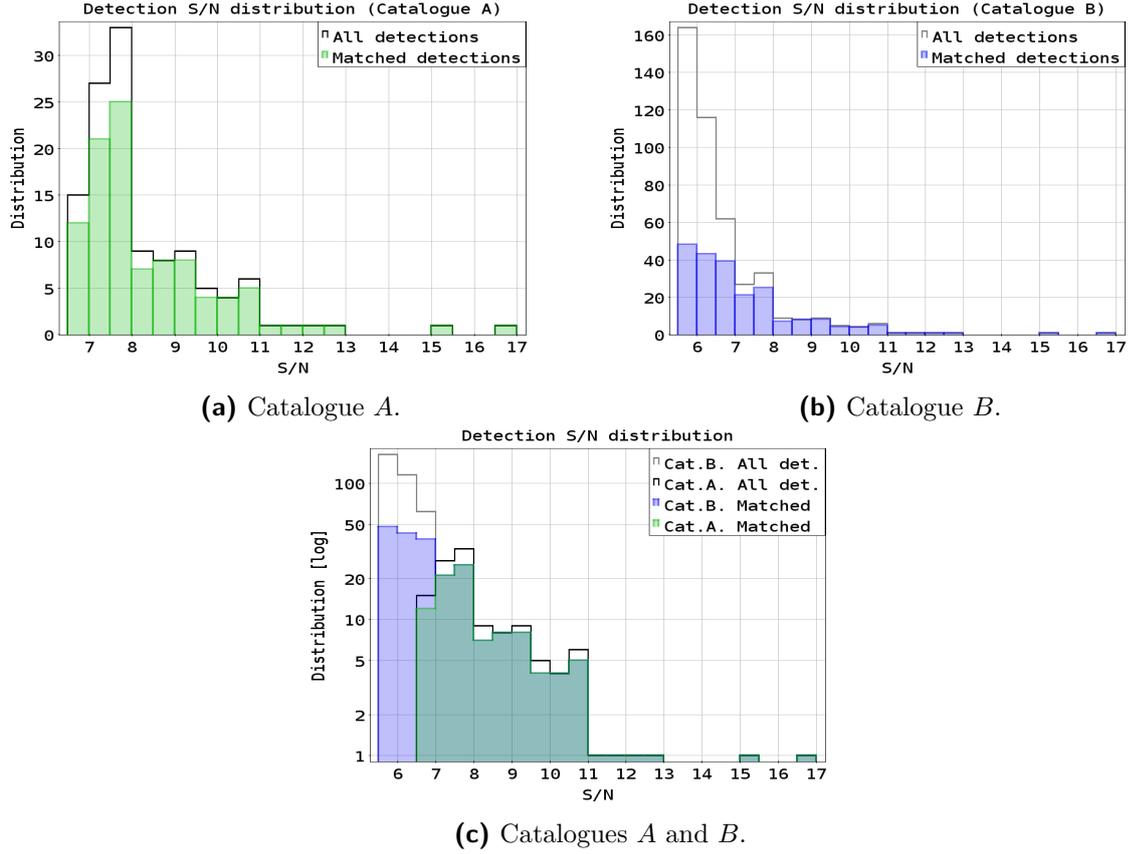
**Figure 4.9:** Distribution of the signal to noise ratio for all the detections and the matched detections of the two catalogues. Catalogue *A* is shown in **panel4.9a**, catalogue *B* is shown in **panel 4.9b**. They are both shown in **panel 4.9c** for a direct comparison. These histograms show graphically that the two catalogues have the same detections and matches starting from the S/N bin $[7.0; 7.5]$. We matched our detections with the haloes of the simulation, that have mass in the range $[0.74, 14.0] \times 10^{14} \, \mathrm{M_\odot} \, h^{-1}$ and redshift in the range $[0, 1]$.

**Figure 4.10:** Percentage of matched detections above a given S/N. For instance, 90.4% of the detections with S/N > 9.50 are matched in both catalogue *A* and *B*.



**(a)** Catalogue *A*.  **(b)** Catalogue *B*.
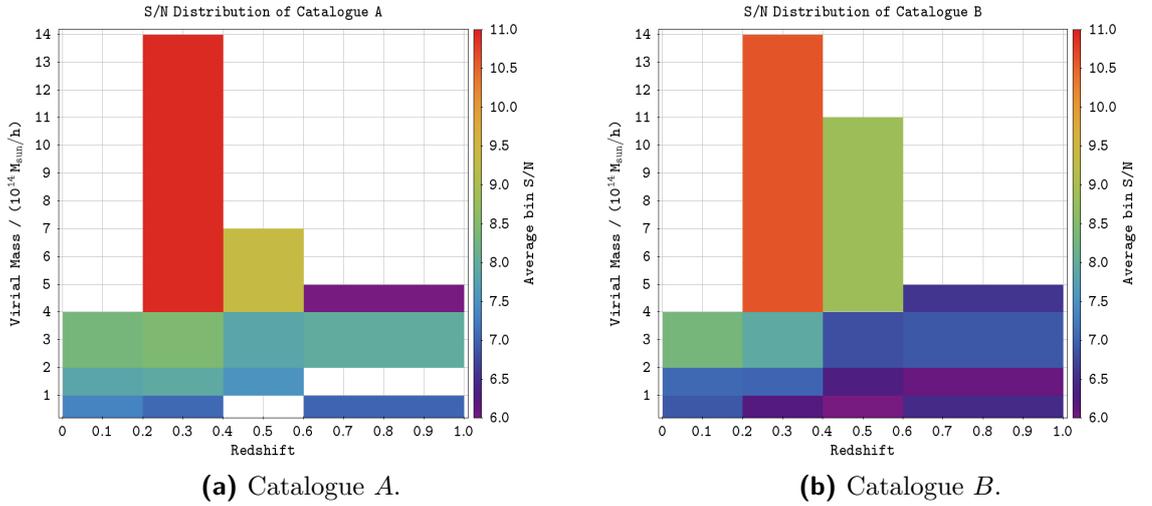
**Figure 4.11:** S/N ratio distribution of the matched detections as a function of mass and redshift. We plot the average S/N of the matched detections of catalogue *A* (**panel 4.11a**) and *B* (**panel 4.11b**) in each bidimensional bin. In both catalogues the average S/N is higher when the redshift is $\in [0.2; 0.4]$ and the mass is $> 4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$.

**Redshift**

The redshift distribution of the matched haloes with respect to that of all the simulated haloes is shown in figure 4.12, while figure 4.13 shows the percentage of matched haloes for each redshift bin of figure 4.12. The median redshift of the matched detections is 0.35 with quartiles $0.26 \div 0.48$ for catalogue $A$ and 0.32 with quartiles $0.25 \div 0.41$ for catalogue $B$, while the median redshift of all the simulated haloes is 0.62 with quartiles $0.46 \div 78$. We find here further proof of a result we first saw during our discussion of the S/N properties at page 74: the filter is not capable to detect the haloes with the same efficiency at any redshift, it selects preferentially the haloes around the lens redshift used to build the filter (0.3 in this case) and those haloes that have an intermediate distance between observer and nearby background sources, condition that maximises the lensing effects. This reflects the dependence of the geometrical lensing strength on the angular-diameter distances between the observer and the lens, the lens and the sources, and the observer and the sources (Pace et al. 2007). The completeness of our catalogues, i.e. the percentage of matched haloes, has a peak extended from redshift $\approx 0.15$ to $\approx 0.35$, as it is shown in figure 4.13. Another observation from this figure is that completeness systematically increases in catalogue $B$; this suggests that lowering the threshold introduces a relevant fraction of reliable detections despite the matching fraction of the overall catalogue decreases.

The filter does not just detect haloes with a privileged redshift, the mass of the halo is also relevant. In particular, we expect the completeness to be related to the mass: at a given redshift, more massive lenses are expected to produce stronger lensing signal. We investigate the interplay between mass and redshift in figure 4.14, that plots the catalogues completeness as a function of redshift, with different cuts in mass. In particular in figure 4.14a we have plotted the completeness of catalogue $A$ as a function of redshift for 4 subset of haloes: haloes with virial mass $< 1 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$, haloes with mass $\in [1;2] \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$, haloes with mass $\in [2;4] \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$ , haloes with mass $> 4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$. We do the same in figure 4.14b for catalogue $B$. Figure 4.14c shows the completeness of catalogue $A$ as a function of redshift for the haloes with 4 different cuts in mass: haloes with mass $> 4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$, haloes with mass $> 2 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$, haloes with mass $> 1 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$, all the haloes. We do the same in figure 4.14d for catalogue $B$. These plots also confirm the existence of a peak extended around redshift $\approx 0.15 \div 0.35$ and that the completeness of catalogue $B$ is higher than catalogue $A$, hence using the weak threshold can actually improve the completeness of our catalogue. They also show that, as a general trend, when the mass increases the matching percentage increases too, although this is not always the case as there are some exceptions. Mainly, the massive haloes $> 4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$ show significant fluctuations due to their small number: at $z < 0.2$ and $z > 0.6$ not every massive halo is detected; we might even detect smaller percentages than the lower mass haloes. A possible explanation lies in the poor statistics of these haloes: they are only 31, so a high background noise fluctuation that

lowers the S/N might be responsible not detecting some massive haloes, with a resulting strong loss of percentage. Let us also note that we match 18 haloes in cat. *A* and 21 in *B* out of the 31. As an alternative, losing massive haloes at low and high redshift might also be a suggestion that some of the lower mass matched detections might actually be random matches between a halo and a spurious detection caused by noise. Figure 4.14 also shows that the peak of completeness around redshift $\approx 0.15 \div 0.35$ is always found, confirming that the efficiency of detecting haloes has a strong dependence on the halo redshift.

The distribution of the matched haloes with respect to all the haloes as a function of both mass and redshift is plotted in figure 4.15, where it becomes clear that most of the matched detections have intermediate redshift $\approx 0.2 \div 0.4$ and intermediate-high mass $> 2 \times 10^{14} \, \mathrm{M}_{\odot} \, h^{-1}$. The distribution in figure 4.15 is turned into maps of completeness in figure 4.16. In order to realise them we divided, for each bidimensional bin, the number of matched haloes in catalogues *A* (figure 4.16a) and *B* (figure 4.16b) by the total number of simulated haloes. We have chosen a not uniform binning in order to visualise the same mass subsets used in figures 4.14a and 4.14b and the same redshift subsets that we are going to use in figure 4.19 later in this section. The bins are divided at masses $[1, 2, 4] \times 10^{14} \, \mathrm{M}_{\odot} \, h^{-1}$ and redshifts $[0.2, 0.4, 0.6]$. These plots show that completeness increases in the region of intermediate redshift $\approx 0.2 \div 0.4$ and intermediate-high mass $> 2 \times 10^{14} \, \mathrm{M}_{\odot} \, h^{-1}$.



**(a)** Catalogue *A*            **(b)** Catalogue *B*
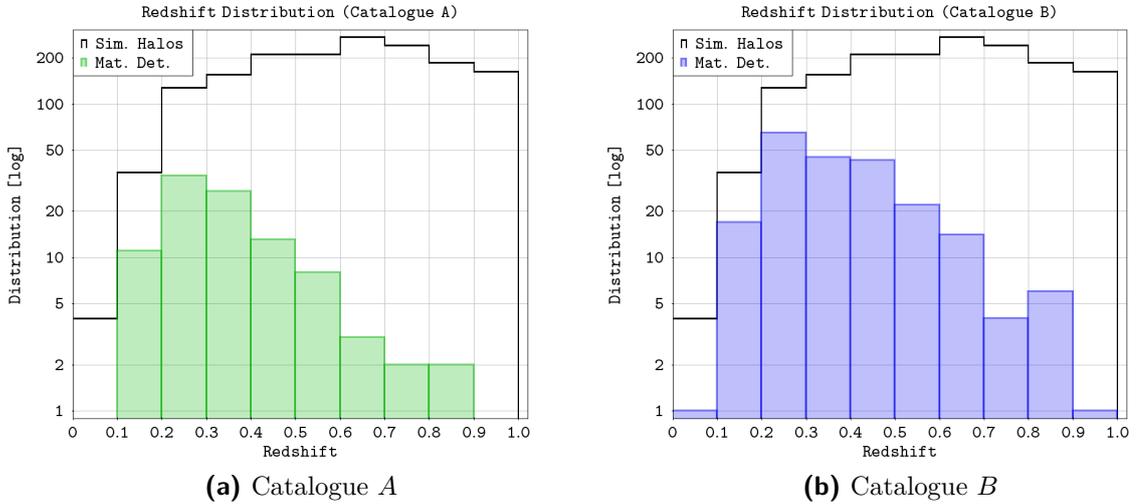
**Figure 4.12:** Redshift distribution of the matched detections and total simulated haloes. We matched our detections with the haloes of the simulation, that have mass in the range $[0.74, 14.0] \times 10^{14} \, \mathrm{M}_{\odot} \, h^{-1}$ and redshift in the range $[0, 1]$.
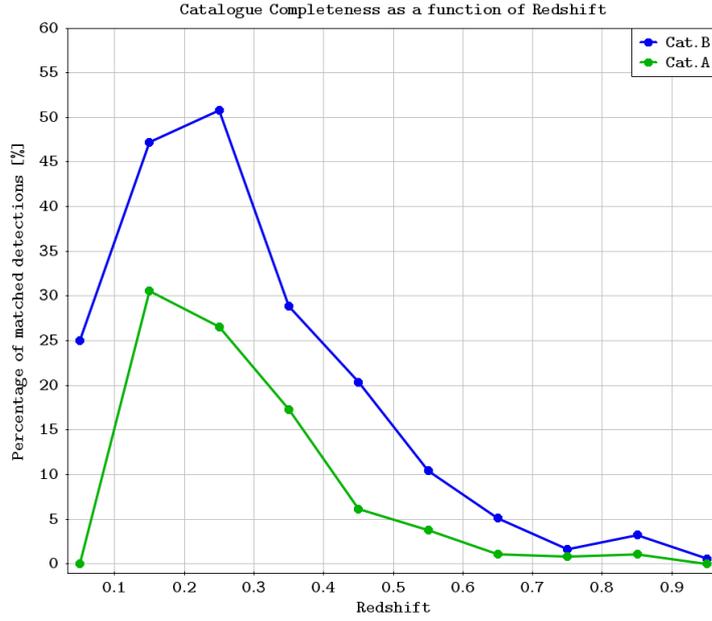
**Figure 4.13:** Catalogue completeness (percentage of matched haloes) as a function of redshift in catalogue *A* (green line) and *B* (blue line). This includes all the haloes of the simulation, that have mass in the range $[0.74, 14.0] \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$. To get these points we divided the number of matched haloes by the total number of haloes in each redshift bin used in figure 4.12, and we placed the points at the center of their bin. There is a clear peak extended between redshift $\approx 0.15$ and $\approx 0.35$ in both the catalogues: the filter does not detect the haloes with the same efficiency at any redshift, but it selects preferentially the haloes around redshift 0.3. The matching percentage increases in catalogue *B*; this suggests that lowering the threshold introduces a relevant fraction of reliable detections.

**(a)** Catalogue *A*, mass subsets.

**(b)** Catalogue *B*, mass subsets.

**(c)** Catalogue *A*, mass cuts.

**(d)** Catalogue *B*, mass cuts.
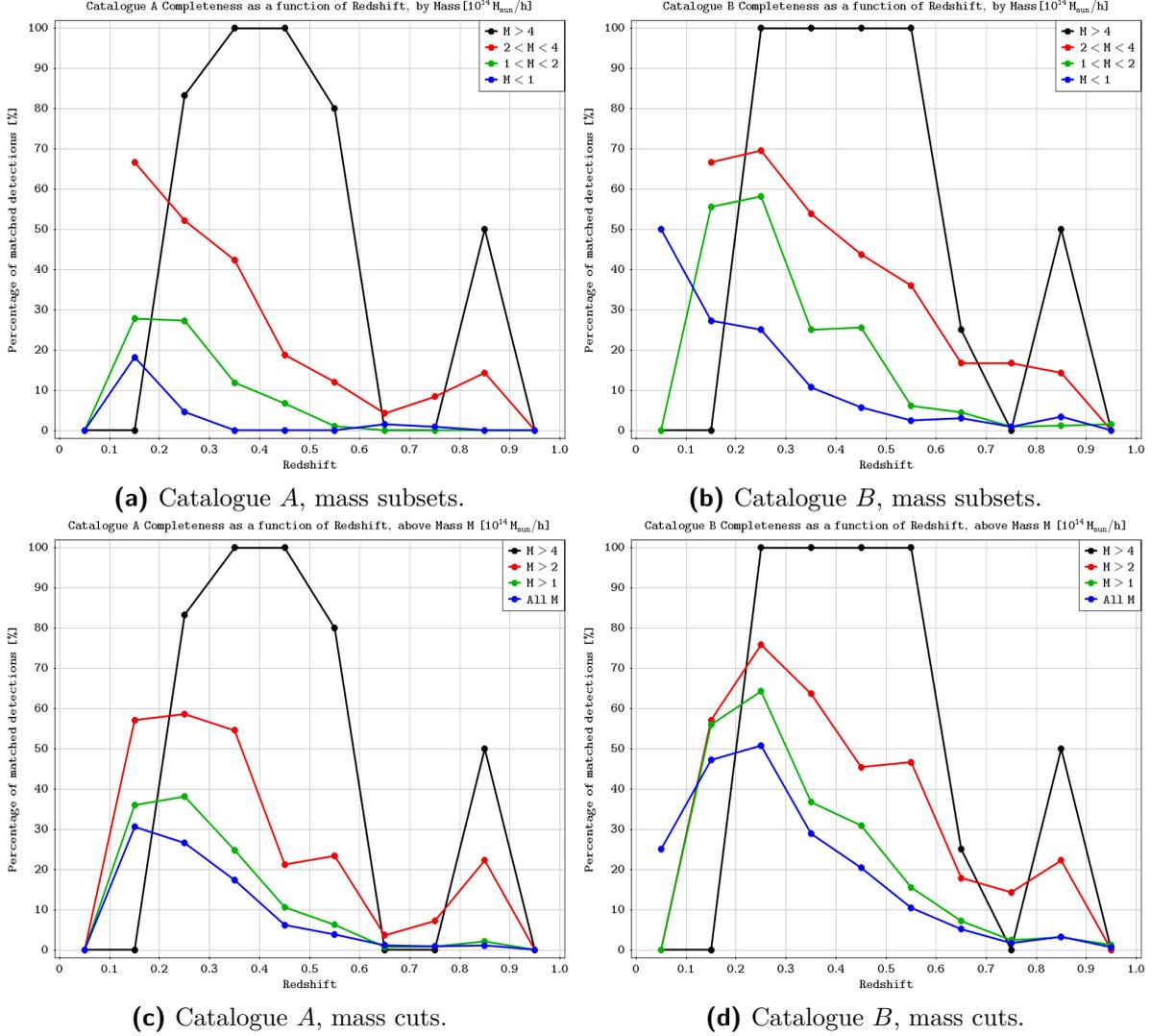
**Figure 4.14:** Catalogue completeness (percentage of matched haloes) as a function of redshift, by virial mass. In **Panels 4.14a** (cat. *A*) and **4.14b** (cat. *B*) we plot the catalogue completeness as a function of redshift for 4 subsets of the halo catalogue according to their mass; in **Panels 4.14c** (cat. *A*) and **4.14d** (cat. *B*) we do the same for 4 subsets with different cuts of mass. As a general trend completeness increases with mass and it is higher in cat. *B* than *A*. The main exceptions are the massive haloes $> 4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$, that show significant fluctuations due to their small number. These plots too show a peak extended around redshift $\approx 0.15 \div 0.35$.
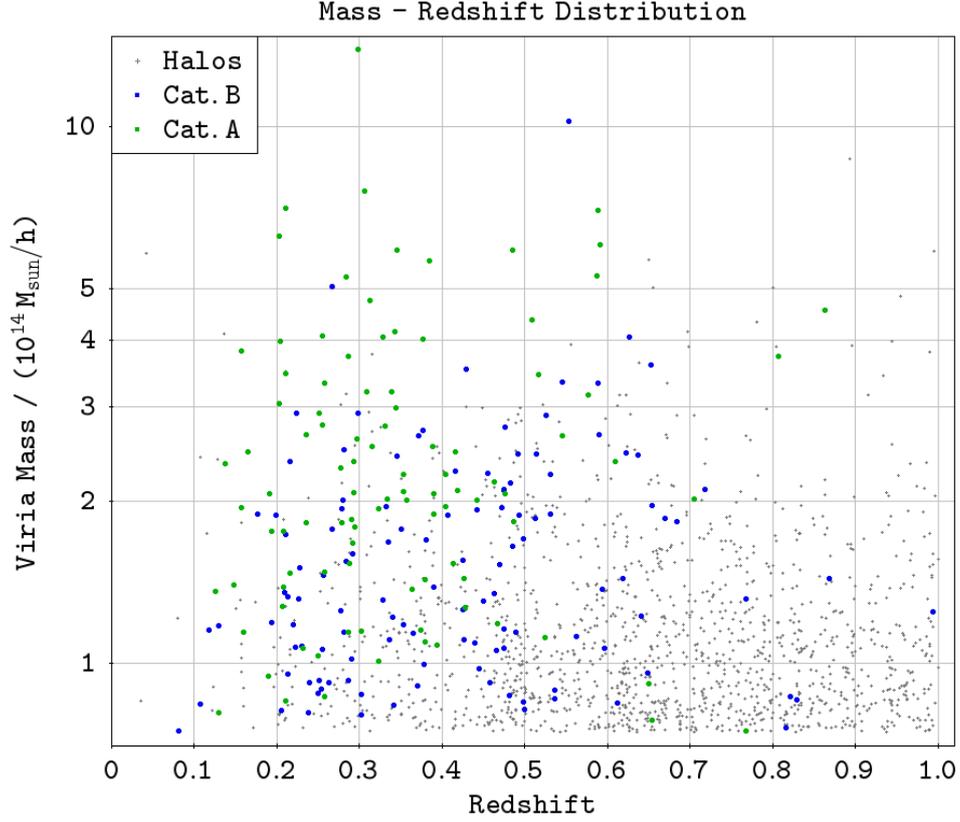
**Figure 4.15:** Distribution of the haloes as a function of Mass and Redshift. Grey crosses are all the simulated haloes. Blue circles are the matched detections of catalogue *B*, superimposed to the cross of their halo. Green circles are the matched detections of catalogue *A*, superimposed to their counterparts in catalogue *B* and halo. The grey crosses still visible are then the haloes not matched with catalogue *B*. Most of the matched haloes have intermediate redshift $\approx 0.2 \div 0.4$ and intermediate-high mass $> 2 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$.

**(a)** Catalogue *A*.
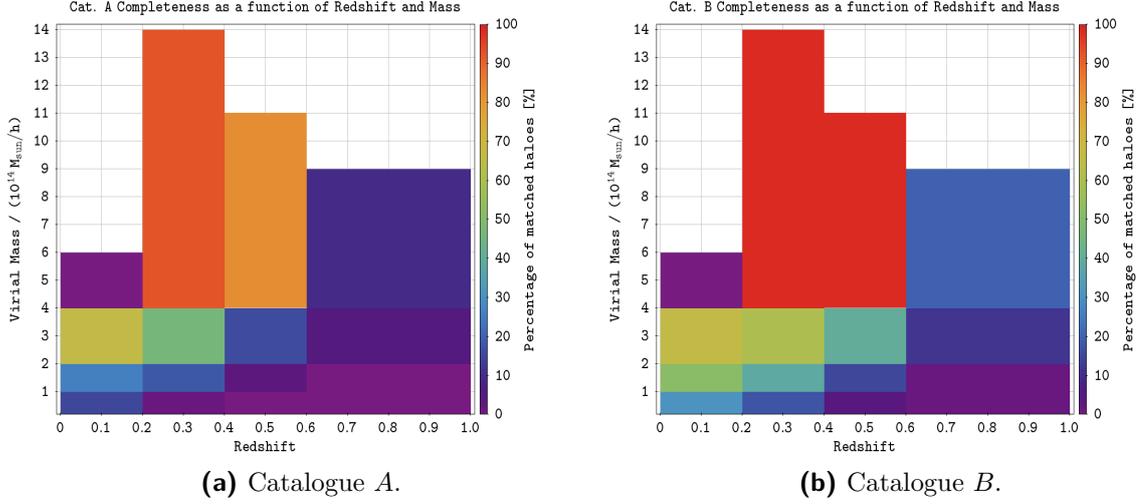
**(b)** Catalogue *B*.

**Figure 4.16:** Percentage of matched haloes (completeness) as a function of mass and redshift for the two catalogues. Each bidimensional bin shows the percentage of matched detections (over all the simulated haloes) in that bin, for cat. *A* in **panels 4.16a** and for cat. *B* in **panels 4.16b**. Most of the haloes with intermediate redshift $\approx 0.2 \div 0.4$ and intermediate-high mass $> 2 \times 10^{14}\,\mathrm{M_\odot}\,h^{-1}$ are matched.

## Virial Mass

The virial mass distribution of the matched detections is shown in figure 4.17, where it is compared with the distribution of all the simulated haloes. These histograms show that not every mass can be detected with the same efficiency: in units of $10^{14}\,\mathrm{M_\odot}\,h^{-1}$ the median mass of the simulated haloes is 1.07 with quartiles $0.86 \div 1.47$, while the median mass of the matched detections is 2.07 with quartiles $1.44 \div 3.33$ in catalogue *A* and 1.78 with quartiles $1.14 \div 2.46$ in catalogue *B*. They also show that not every massive halo produces a lensing signal so strong to be always detected: the second most massive haloes of the simulation (halo $H\#12$) is detected and matched only in catalogue *B*; the third most massive (halo $H\#1$) is not detected in either of them. They are described in table 4.2. This can be probably attributed to their high redshift: halo $H\#12$ has redshift 0.55, near the peak of completeness, halo $H\#1$ has redshift 0.89, far away from the peak. The detection of $H\#12$ only in catalogue *B* is one example of the perks of using the weak threshold: it introduces many spurious detections, but increases the number of detection and matches at high redshift, as it was shown in figure 4.13.

Figure 4.18 describes catalogues *A* and *B* completeness, i.e. the percentage of matched haloes, as a function of the virial mass of the haloes. The points shown in figure 4.18a are found by dividing the number of matched haloes in the mass bins of figure 4.17 by the number of all the simulated haloes in that bin. A "cumulative" version of figure 4.18a can be seen in figure 4.18b, showing the percentage of matched detections with

**Table 4.2:** Second and third most massive haloes of the simulation. The second is detected only in catalogue $B$, the third in neither of them. We report their ID in the simulated halo catalogue, their position in it, mass, redshift and the ID and S/N of the detection of catalogue $B$ with which they have been matched.

| ID [sim.] | RA [deg] | DEC [deg] | Mass $\left[M_\odot\, h^{-1}\right]$ | Redshift | ID [B] | S/N |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| $H\#12$ | 2.9990 | 0.2390 | $1.02 \times 10^{15}$ | 0.55 | $B\#257$ | 6.393 |
| $H\#1$ | 1.5005 | -3.9082 | $8.71 \times 10^{14}$ | 0.89 | - | - |

virial mass above a given value. At $4 \times 10^{14}\, M_\odot\, h^{-1}$ we reach 68% of completeness for catalogue $B$. The percentages are steadily growing functions of the mass until we get to $5 \times 10^{14}\, M_\odot\, h^{-1}$; after that the poor statistics makes them subject to strong fluctuations. Those are responsible for the dips in completeness that can be seen at $(7 \div 8) \times 10^{14}\, M_\odot\, h^{-1}$: there are only 5 haloes in total with mass $> 7 \times 10^{14}\, M_\odot\, h^{-1}$; one of them, the second most massive halo, is detected only in $A$. The other one, the third most massive halo, is lost in both catalogues because of its high redshift. The completeness is lower for catalogue $A$, another confirmation that using the weak threshold can in fact improve the catalogue completeness.

We have already seen that the redshift is a crucial quantity in the detection process of halo: we expect the completeness to be higher at intermediate redshift. For this reason we have plotted in figure 4.19 the completeness of our catalogues as a function of mass or above a given mass after we had divided the haloes in four subsets according to their redshift: $z < 0.2$, $0.2 < z < 0.4$, $0.4 < z < 0.6$, $z > 0.6$. In particular we plot the completeness as a function of mass in figures 4.19a (catalogue $A$) and 4.19b (catalogue $B$); we have used the same binning of figure 4.18 since we are basically splitting its plots into three redshift subsets. We plot completeness above a given virial mass in figures 4.19c (cat. $A$) and 4.19d (cat. $B$). For $z < 0.2$ and $z > 0.6$ our filter does not find many haloes, hence the completeness above a given mass never goes above 55%. On the contrary, for haloes with $0.2 < z < 0.4$ the matching percentage above a given mass increases when the mass increases, is always above $\approx 20\%$ for cat. $A$ and $\approx 40\%$ for cat. $B$ and it reaches 100% completeness at mass $5 \times 10^{14}\, M_\odot\, h^{-1}$ (cat. $B$) and $6 \times 10^{14}\, M_\odot\, h^{-1}$ (cat. $A$). For haloes with $0.4 < z < 0.6$ the matching percentage above a given mass increases when the mass increases, but at $4 \times 10^{14}\, M_\odot\, h^{-1}$ its behaviour is different in the two catalogues: it decreases in cat. $A$ and it reaches and stays at 100% completeness in $B$. We thus once again conclude that the filter is optimised for searching haloes at $0.2 < z < 0.4$. The filter can still detect the majority of haloes with $0.4 < z < 0.6$, especially if they are not too small ($> 3 \times 10^{14}\, M_\odot\, h^{-1}$). The filter becomes less effective in detecting haloes at $z > 0.6$; fluctuations in the curves are mostly due to small statistics.
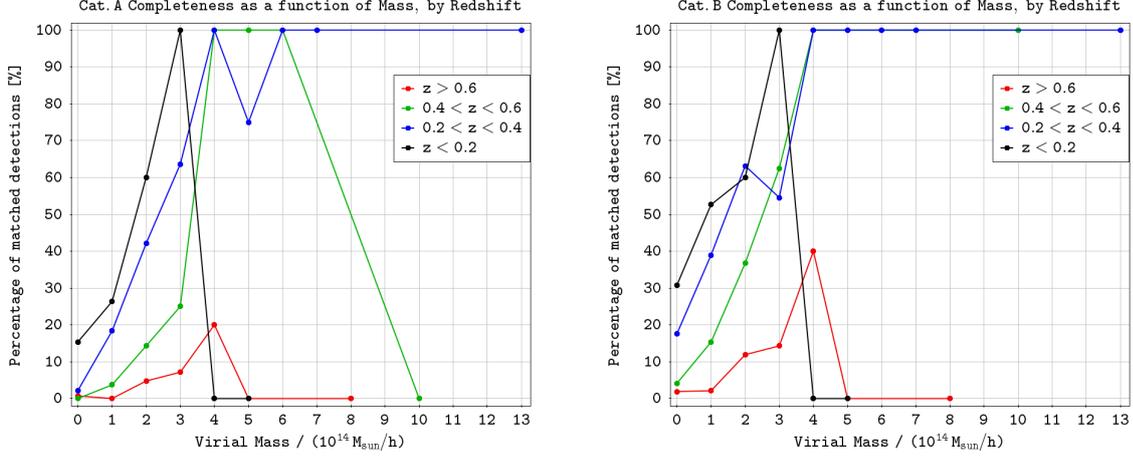
**(a)** Catalogue $A$  |  **(b)** Catalogue $B$
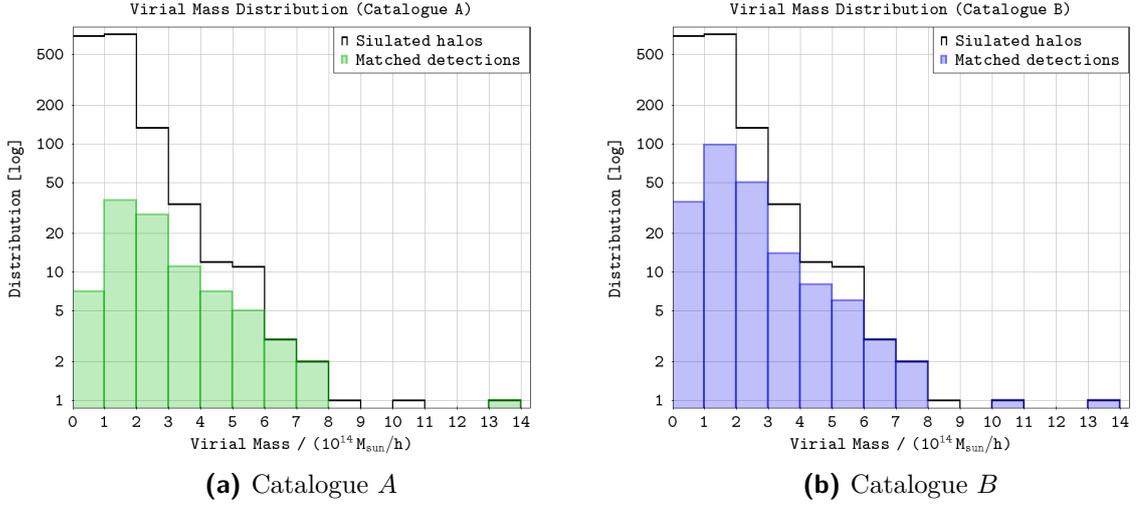
**Figure 4.17:** Distribution of the virial mass $M_{200}$ for the matched haloes of catalogue $A$ (**panel 4.17a**), $B$ (**panel 4.17b**), and all the simulated haloes, that have redshift $\in [0, 1]$. Note that the second most massive halo is detected and matched only in catalogue $B$, while the third is lost in both catalogues. This can be probably attributed to their high redshift (0.55 and 0.89 respectively) that causes them to have a lower S/N.
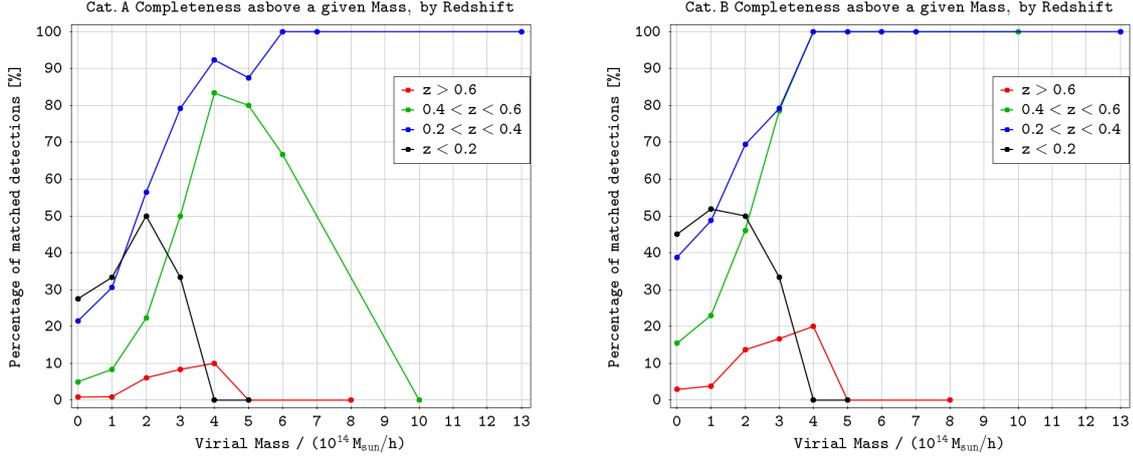


**(a)** Percentage of matched haloes as a function of the virial mass.  |  **(b)** Percentage of matched haloes above a given virial mass.
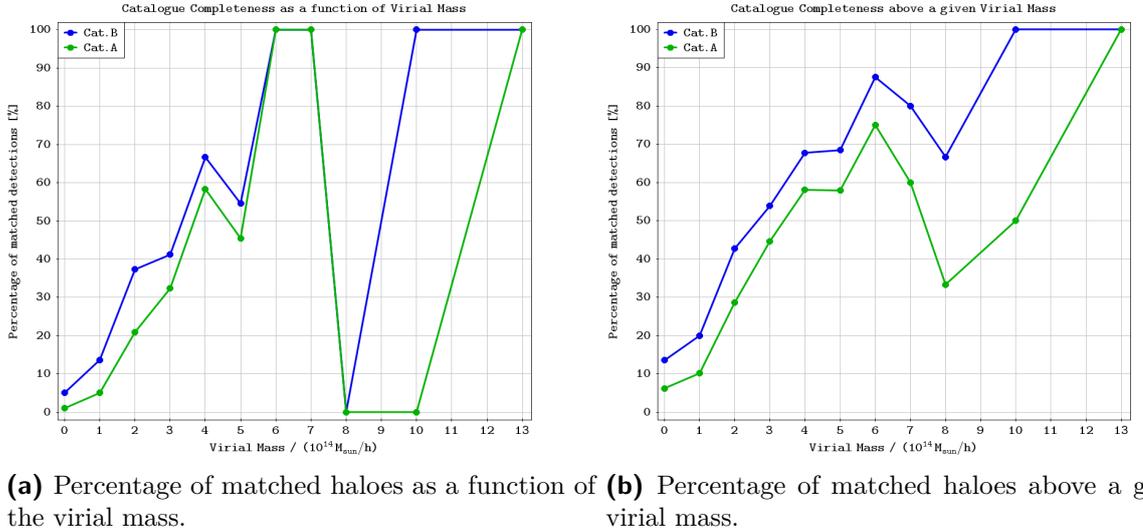
**Figure 4.18:** Percentage of matched haloes as a function of the virial mass (**panel 4.18a**) and above a given virial mass (**panel 4.18b**), for all the redshift of the haloes ($z \in [0, 1]$). The points are placed at the lower end of the bin used to compute them, as discussed in section 4.2.2. Both graphs show that the catalogue completeness increases when the mass increases. The percentages at masses higher than $> 7 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$ are computed on a very small number of haloes, hence they are subject to strong fluctuations and a dip in percentages due to losing the third most massive halo of the simulation (and the second too in catalogue $A$).

**(a)** Percentage of matched haloes as a function of the virial mass, by redshift, catalogue *A*.

**(b)** Percentage of matched haloes as a function of the virial mass, by redshift, catalogue *B*.

**(c)** Percentage of matched haloes above a given virial mass, by redshift, catalogue *A*.

**(d)** Percentage of matched haloes above a given virial mass, by redshift, catalogue *B*.

**Figure 4.19:** Percentage of matched haloes as a function of the virial mass, by redshift. These graphs show how the completeness of the catalogues changes as a function of halo redshift. The sources with redshift $z \in [0.2; 0.6]$ have the highest matching percentages; this is caused by the peak in detection efficiency of the filter as a function of redshift. Another notable trend is that the matching percentage tends to increase as a function of the halo mass, although there are fluctuations given by the poor statistics.

## 4.3 Application of a "Customised" filter on the catalogue

We analysed again the data by using an optimal filter with the following parameters:

- Lens mass: $1.0 \times 10^{15} \, \mathrm{M}_\odot \, h^{-1}$

- Lens redshift: 0.60

- Source redshift: 0.83.

- White galaxy noise: $1.88 \times 10^{-10} \, \mathrm{rad}^2$.

- Beam FWHM: $0.18'$.

This is a set of parameters "customised" on the properties of our catalogue of simulated ellipticities: the lens and source redshift are set exactly equal to the median values respectively of the haloes and galaxies redshift distributions, the beam FWHM is the average galaxy distance and the level of the white galaxy noise is *exactly* equal to the one of the catalogue, as described in section 4.1. This filter has cut radius $45.33'$, that GtH21 once again extends to $60'$. The average radius of this filter is similar to the one of the "standard" filter, $2.19'$, so each pixel is large $1/5$ of this scale, i.e. $0.44'$, as for the "standard" filter. We show the customised filter used in figure 4.20, in the real space in figure 4.20a and in the Fourier space in figure 4.20b; we also show the standard filter of section 4.2 for comparison. The two filters are very similar, so we expect the results of the customised filter to be similar to the ones of the standard filter. Their main difference is the redshift of the lens: by using the customised filter we hope to improve the detection of haloes with redshift $\approx 0.6$. We decided to change the lens redshift because Maturi et al. (2007) had already shown that the optimal filter was stable against changes of the filter size given by the halo mass, due to the shape control of the filter imposed by the noise power spectrum. We run GtH21 and analysed the results exactly as we did in section 4.2. We report in this section the main difference with those results; full analysis is reported in appendix B.

After running GtH21 we studied the distribution of the S/N of the T-map and R-map and we computed the strong and weak S/N thresholds above which a pixel is assumed to contain a detection as described in section 3.2.4 and as we already did in section 4.2. We obtained 7.036 as the strong threshold and 5.614 as the weak threshold. We used them to build two catalogues: **catalogue C** is the one built with the strong threshold and it is made up by 96 detections; **catalogue D** the one built with the weak threshold and it is made up by 426 detections. Every detection in $C$ has an equivalent detection in $D$, with the same position and S/N but a smaller area and a different ID number. We matched our detections in catalogues $C$ and $D$ with the simulated haloes by using *TOPCAT*

Sky algorithm once again, with a maximum mismatch distance of 1.5′. When using this maximum mismatch distance *TOPCAT* matched 81 out of 96 detections (84.4%) of catalogue $C$ and 221 out of 426 detections (51.9%) of catalogue $D$. All the halos matched with a detection in catalogue $C$ have been matched with the equivalent detection in $D$, except for one: halo $H\#861$ has been matched to detections $C\#23$ and $D\#99$, despite $C\#23$ being the equivalent of $D\#100$. We believe that $D\#99$, a single-pixel detection, is either a weaker part of detection $D\#100$ that has been fragmented due to noise, or a spurious peak that happens to arise between $D\#100$ and $H\#861$. See appendix B for a more accurate discussion of these detections.
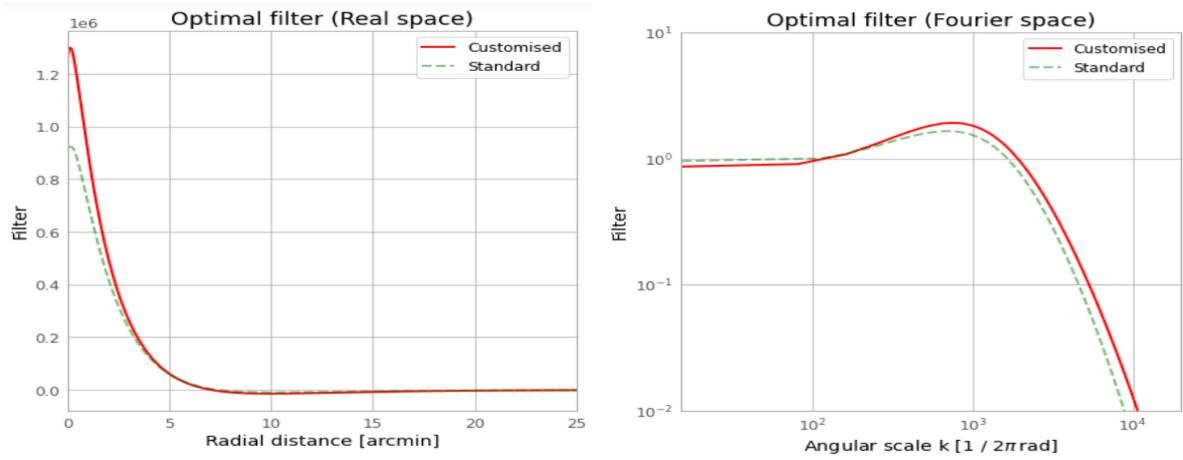


**(a)** "Customised" optimal filter in real domain.  **(b)** "Customised" optimal filter in the Fourier space.

**Figure 4.20:** Application of GtH21 on the data. The "Customised" optimal filter applied on the data and discussed in section 4.3 is shown in the real space in **panel 4.20a**, and in the Fourier space in **panel 4.20b**, as red lines. The green dashed curves show the "Standard" filter discussed in section 4.2 for comparison. The "Customised" optimal filter is very similar to the Standard filter, their main difference is the higher central normalization in the real space. We thus expect the results of the customised filter to be similar to the ones of the standard filter.

In figure 4.21 we show the percentage of matched detections with S/N above a given value. In order to plot them we computed the total number of detections and number of matched detections *above* a given S/N in each of our catalogues and divided them. For S/N> 7.0 the percentage of matched detections is always above 85% for catalogue $C$, so this can be considered as a reliable S/N value for the construction of a catalogue, confirming the results of Pace et al. (2007). The curve of catalogue $D$ is lower because of source $D\#100$, so we consider the curve of catalogue $C$ to better represent our results after S/N= 7.5. A purity of 100% is reached for sources with S/N > 10.5. There are no strong differences between the results of the two filters: for both filters purity increases as a function of S/N and reaches 100% at S/N≈ 10.5 ÷ 11.
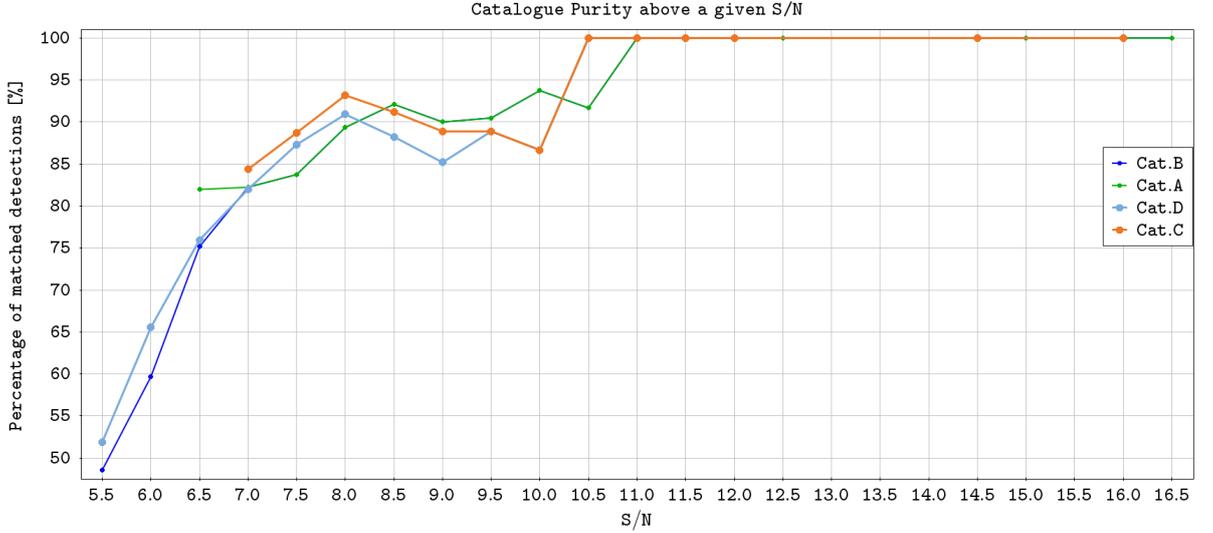
**Figure 4.21:** Percentage of matched detections with S/N above a given value. We computed the total number of detections and number of matched detections *above* a given S/N in each of our catalogues and divided them. Orange is cat. *C*, light blue is cat. *D*; we also plotted a green line for cat. *A* and a dark blue line for cat. *B* for comparison. After S/N= 7.0 the dark blue curve of cat. *B* is equal to the green curve of cat. *A*, superimposed onto it. The same is true for catalogues *C* and *D* after S/N= 9.5.

We plot the completeness of our catalogues, i.e. the percentage of matched haloes, as a function of redshift in figure 4.22. As we found for catalogues *A* and *B* in section 4.2.2 there is a clear peak of completeness extended from redshift $\approx 0.15$ to $\approx 0.35$, and completeness systematically increases in the catalogue with lower threshold (cat.*D*). The filter preferentially selects haloes at intermediate distances between the observer and the sources, because of the dependence of the geometrical lensing strength on the angular-diameter distances between the observer and the lens, the lens and the sources, and the observer and the sources (Pace et al. 2007). This can also be seen if we look at the median redshifts of the matched halos: for catalogue *C* it is 0.33 with quartiles $0.25 \div 0.40$, for catalogue *D* it is 0.35 with quartiles $0.26 \div 0.49$, for all the simulated haloes is 0.62 with quartiles $0.46 \div 0.78$. Although we expected the customised filter to perform better than the standard one at redshift $\approx 0.6$ this is not the case: there are no strong differences between the results of the two filters. This can be attributed to the fact that the weak lensing effects of haloes at $z \simeq 0.6$ on the background galaxies are not intrinsically strong as they would be if they were at $z \simeq 0.3$, where we find most of the haloes we detect. Let us note that the curve for cat. *C* is systematically lower than the one for cat. *A*; this might be attributed to the fact that *C* was built with a more restrictive threshold than *A* and it thus it is able to detect and consequently match less haloes than *A*.
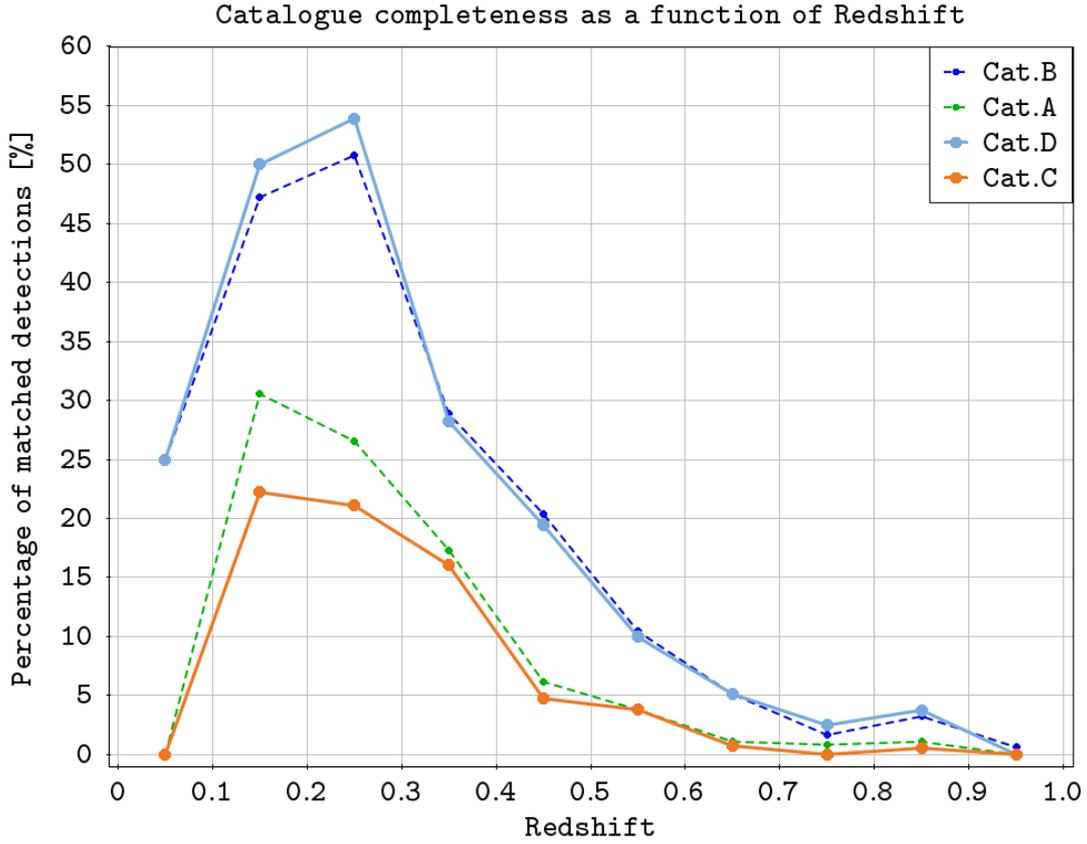
**Figure 4.22:** Percentage of matched haloes as a function of redshift. In orange cat. *C*, in light blue cat. *D*; we also plotted a green line for cat. *A* and a dark blue line for cat. *B* for comparison. The most notable feature of the plot is the peak of completeness extended from redshift $\approx 0.15$ to $\approx 0.35$.

The completeness of the catalogues as a function of mass is shown in figure 4.23, where we plot the percentage of matched detections with virial mass above a given value. The percentages are steadily growing functions of the mass until $4 \times 10^{14} \, \mathrm{M_\odot} \, h^{-1}$, where catalogue $D$ reaches 65% completeness; after that they are subject to strong fluctuations due to the small number of the high-mass haloes. There are in fact only 5 haloes in total with mass $> 7 \times 10^{14} \, \mathrm{M_\odot} \, h^{-1}$; catalogue $C$ does not detect three of them ($H\#12$, $H\#1$, $H\#194$) while $D$ does not detect $H\#1$. These haloes are not matched probably because of their high redshift, 0.55 for $H\#12$ and 0.89 for $H\#1$, and the more restrictive S/N threshold of catalogue $D$ for $H\#194$. Those are responsible for the dips in completeness that can be seen at high mass values. This prevents the completeness functions from reaching 100% but in the last bin (due to the detection of the most massive halo of the simulation in both catalogues). The completeness is lower for catalogue $C$, another confirmation that lowering the threshold can in fact improve the catalogue completeness. The curves of catalogues $B$ and $D$ do not show notable differences; the curve of catalogue $C$ is systematically lower than catalogue $A$. This might be attributed to the higher threshold of cat. $C$.
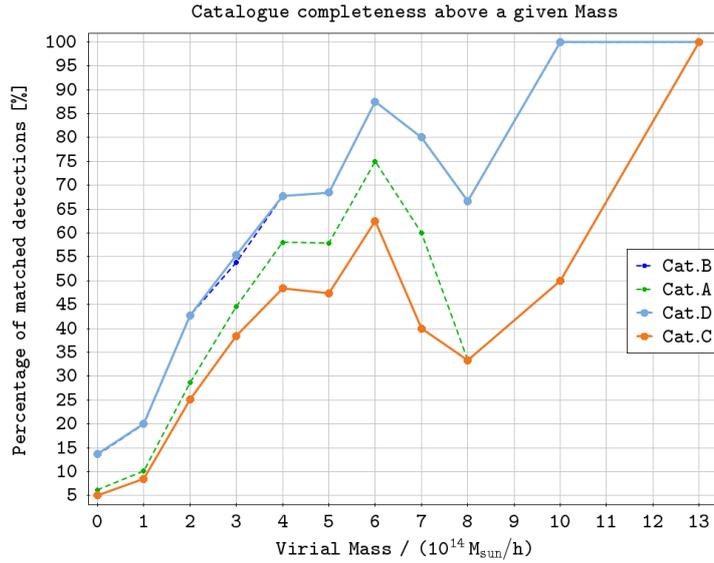


**Figure 4.23:** Percentage of matched haloes with virial mass above a given virial mass. In orange cat. $C$, in light blue cat. $D$; we also plotted a green line for cat. $A$ and a dark blue line for cat. $B$ for comparison. The percentages at higher masses than $> 7 \times 10^{14} \, \mathrm{M_\odot} \, h^{-1}$ are computed on a very small number of haloes, hence they are subject to strong fluctuations and a dip in percentages due to losing some of the most massive haloes of the simulation,

The analysis we carried out in the current section showed that, although we expected the customised filter to perform better than the standard one at redshift $\geq 0.5$ this is not the case; we were not able to detect significant differences between the results of the two

filters that could be attributed to changing the lens redshift. This probably happened because the weak lensing effects of haloes at $z \simeq 0.6$ on the background galaxies are not intrinsically strong as those of the haloes at $z \simeq 0.3$, where we find most of the haloes we detect, even for a catalogue of ellipticities deep as ours and representative of the weak lensing data that Euclid is going to measure.

# Conclusion and future perspectives

We developed a new version of the *Get the Halo* code realized more than fifteen years ago by Matteo Maturi, *Get the Halo* 2021, that implements the "optimal" linear matched filter defined by Maturi et al. (2005). Our code is compatible with the 2017 standard of the `C++` language and it replaces the *astrocpp* library with the *CosmoBolognalib* libraries (Marulli et al. 2016). We applied it to a catalogue of simulated galaxy ellipticities weak-lensed by a catalogue of simulated halos. They were simulated to represent typical data that will be available thanks to the *Euclid* mission, hence we our analysis can forecast the filter performances on the weak lensing data that are going to be measured by *Euclid*.

We built two main catalogues of detections, catalogues $A$ and $B$ in section 4.3, that we matched with the catalogue of simulated halos in order to evaluate their purity and completeness. Our results confirms the expectations on the filter performance raised by Maturi et al. (2005) and Pace et al. (2007).

The first prediction we were able to confirm is that a value of the S/N$\approx 7$ can be considered as a reliable threshold to detect haloes through weak lensing. For instance, we noted that $\simeq 83\%$ of our detections with S/N$> 7$ were matched. The remaining 17% might instead be spurious detections caused by weak lensing from the large-scale structure. The purity of our catalogues increases as a function of S/N and reaches 100% at S/N $\approx 10.5 \div 11$.

Another relevant prediction that we were able to confirm is that the filter selects preferentially the haloes with redshift between 0.2 and 0.5, that have an intermediate distance between observer and nearby background sources, condition that maximises the lensing effects. This reflects the dependence of the geometrical lensing strength on the angular-diameter distances between the observer and the lens, the lens and the sources, and the observer and the sources. We saw this several times during our analysis, for instance when we noted that the average S/N of the matched detections is higher when their redshift is $\in [0.2, 0.4]$ and their mass is $2 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$. A consequence of this *selection* effect of the filter was that the completeness of our catalogues always showed a peak for redshifts approximately between 0.2 and 0.4. The completeness function is related to the mass of the matched haloes: we found that the percentage of matched haloes above a given mass is a steadily growing function of the mass until we reach $(4 \div 5) \, \mathrm{M}_\odot \, h^{-1}$, where it reaches values $\approx 68\%$ for catalogue $B$ and $\approx 58\%$ for catalogue

*A*. After $(4 \div 5)\,\mathrm{M}_{\odot}\,h^{-1}$ the small number of haloes in our simulations causes strong fluctuations in the completeness.

We applied again the filter initialised with a different set of parameters (the main parameter we changed was the lens redshift, increased from 0.3 to 0.6), and we built two other catalogues of detections, catalogues *C* and *D* in section 4.3. Although we hoped to improve the filter performances at $z \geq 0.5$, we did not find relevant differences with the first filter. This can be attributed to the fact that the weak lensing effects of haloes at $z \simeq 0.6$ on the background galaxies are not intrinsically strong as those of the haloes at $z \simeq 0.3$, even for data as deep as the ones we analysed, representative of the weak lensing data that Euclid is going to measure. We were still able to confirm the expectations of Maturi et al. (2005) and Pace et al. (2007).

Our work is part of the "test challenge" organised inside the *Euclid SWG Clusters of Galaxies* and more specifically as a part of the activities of its WP 10, "Weak Lensing Selected Clusters". The goal of the challenge is to select the best cluster detection algorithm that uses weak lensing data, by testing several detection methods on the same dataset. The challenge is going to take place during the rest of 2021 and into 2022. We are still in the early phases of the challenge, hence we were allowed to match our detections with the simulated haloes; this will not be allowed during later phases, when the algorithms will be subject to *blind* testing. Our results are the very first to be produced and presented, so the other participants of the challenge will need to compare their algorithms and results to the ones shown in this thesis. The best algorithm might be used to detect galaxy clusters as a complementary algorithm to *AMICO* and *PZWav* in the Euclid data analysis pipeline, chosen to identify galaxy clusters in photometric data (Euclid Collaboration et al. 2019). Our code has an intrinsic advantage over any other: it is a *optimal linear matched filter* just like *AMICO* (Bellagamba et al. 2018); their difference is that *AMICO* uses photometric data while *GtH*21 uses weak lensing catalogues of galaxy ellipticities. Since they have the same properties they can be used together when both photometric and weak lensing data are available: their likelihoods can be summed (Bellagamba et al. 2018) so that the S/N of the common detections increases. Using them together might then have two consequences: improving the significance of the detections of *AMICO* and raise the total number of detections, because the S/N of some detections might increase over the minimum threshold required. The final objective of our algorithm GtH21 might then be to enhance the reliability of *AMICO* and the entire *Euclid* data analysis pipeline.

# Acknowledgements

# Ringraziamenti

Concludere un lavoro di tesi con dei ringraziamenti può forse apparire scontato, ma la gratitudine non è mai abbastanza. Per questo vorrei esprimere anche in italiano alcuni ringraziamenti ai docenti che mi hanno seguito in questi mesi.

Il primo è per il mio relatore, il Prof. Lauro Moscardini. Non sarei riuscito a realizzare questo lavoro di tesi senza il suo supporto, la sua disponibilità e prontezza nel rispondere sempre ad ogni mio dubbio, gli spunti di riflessione offerti e l'inesauribile pazienza dimostrata nei miei confronti. I suoi consigli e la sua fiducia in tutti questi mesi sono stati un dono insperato.

Sono estremamente grato al mio correlatore, il Prof. Matteo Maturi, per aver messo a mia disposizione il suo codice *Get the Halo* e per il suo aiuto e i suoi consigli durante tutto il lavoro di tesi, dall'installazione della libreria *astrocpp* alla revisione finale.

Altro sentito ringraziamento va al Prof. Carlo Giocoli per aver messo a mia disposizione il suo codice per il calcolo dello spettro di potenza dello shear cosmico e per essere stato sempre disponibile e pronto ad aiutarmi.

La vostra competenza e comprensione saranno un esempio per il mio futuro.

Desidero essere, almeno questa volta, non poco sentimentale e ringraziare tutti gli amici che, dovunque si trovassero, mi hanno sempre supportato.

Un profondo grazie va ai ragazzi che ho conosciuto a Bologna, sia in Camplus che in facoltà, per avermi accompagnato in questi cinque anni ed averli resi speciali.

Grazie a Stefano, brillante futuro astrofisico, per il sostegno e le lunghe chiacchierate sulle nostre vite e i nostri percorsi, contemporaneamente paralleli e strettamente intrecciati.

Un ringraziamento altrettanto profondo va agli amici della comitiva con cui sono cresciuto, per il loro affetto che mai si è affievolito e il legame che continuiamo a condividere nonostante viviamo tutti in città diverse.

Desidero infine dedicare un ringraziamento e un pensiero ai miei familiari.

Devo loro tutto.

Ai miei genitori, Evelina e Carmelo, che hanno sempre posto me prima di ogni loro necessità, hanno sempre supportato le mie scelte e mi hanno fatto *crescere*. Persone come loro sono estremamente rare.

A Federica, la cui energia e vitalità sono eguagliate soltanto dalle stelle più brillanti, per aver arricchito la mia vita come nessun altro.

Alla Nonna, che continua ad essere la persona più forte che conosca e una miniera di preziosi insegnamenti, e a Domenico, fonte inesauribile di gioia e ottimismo, specialmente negli ultimi mesi.

A zia, cugini e al resto della famiglia per il loro affetto e sostegno.

Al Nonno, Salvatore Puglisi, per aver riempito le nostre vite con la sua forza inarrestabile, la sua classe, la sua instancabilità e la sua infinità bontà. A lui dedico questa tesi per essere sempre stato, e continuare ad essere, la mia luce e la mia guida.

# Appendix A

# Setup parameters of Get the Halo 2021

We report here the full list of the parameters that must be given to GtH21 through a single *setup file*, as described in section 3.2.1. The setup file must be a text file where each line is structured according to example (3.17). A typical setup file contains the following parameters:

- **DATA_FILE_IN**: directory and name of the file that contains the input data, it is the first input file discussed above. It is an ASCII file with *HEAD_LINES* lines of header, and then a table of weak lensing data. This is mandatory, with no default value.

- **HEAD_LINES**: number of header lines to be skipped when reading the file. This is required by the *CosmoBolognalib* libraries that read the data from the input file and store them in an object of the class *cbl::data::Table*. This is mandatory, with no default value.

- **FLG_X**: number of the column that contains the first sky coordinate, in arbitrary units. This is mandatory, with no default value.

- **FLG_Y**: number of the column that contains the second sky coordinate, in arbitrary units. This is mandatory, with no default value.

- **FLG_G1**: number of the column that contains the first ellipticity component, adimensional. This is mandatory, with no default value.

- **FLG_G2**: number of the column that contains the second ellipticity component, adimensional. This is mandatory, with no default value.

- **FLG_W**: number of the column that contains the ellipticity weight. This is optional, to ignore set to $-1$ (default) or remove the parameter line.

- **FLG_ZS**: number of the column that contains the redshift. This is optional, to ignore set to −1 (default) or remove the parameter line.

- **FLG_M**: number of the column that contains the shear correction bias. This is optional, to ignore set to −1 (default) or remove the parameter line.

- **X_FAC**: the program works in arcminutes, so this is the factor that converts the first sky coordinate from the units used in the column FLG_X into arcminutes (for instance 60.0 if given in degrees). This is optional, to ignore set to 1.0 (default) or remove the parameter line.

- **Y_FAC**: The program works in arcminutes, so this is the factor that converts the second sky coordinate from the units used in the column FLG_Y into arcminutes (for instance 60.0 if given in degrees). This is optional, to ignore set to 1.0 (default) or remove the parameter line.

- **X_IS_RA**: Usually the first sky coordinate is the right ascension RA, the second is the declination DEC. If the first is RA this must be set to 1, so that the program converts RA into angular degrees on a flat sky approximation. This is optional, to ignore set to 0 (default) or remove the parameter line.

- **DIR_OUT**: Directory where we want to create the output maps file. This is optional, to ignore set to "./" (default) or remove the parameter line.

- **NAME_OUT**: name of the output file.

- **NX**: Number of x-pixels that compose the output map. This is optional, to ignore set to 0 (default) or remove the parameter line; in this case it will be automatically determined. The number of y-pixels will be determined by GtH21 so that the ratio between the numbers of x and y pixels is equal to the ratio between the field of view along x and y.

- **KER_FILE**: directory and name of the file that contains the filter function in the real domain. It is an ASCII file with HE_LI_KER lines of header, then a table of data. Each row is a point where we evaluate the filter function; the column features are the radial distance from zero at which the filter is evaluated and the value of the function in that point. This is mandatory, with no default value.

- **COL_KER_R**: number of the column of the radial position, in arcminutes. This is mandatory, with no default value.

- **COL_KER**: number of column of the filter value. This is mandatory, with no default value.

- **HE_LI_KER**: number of header lines to be skipped when reading the file with the *CosmoBolognalib* class *cbl::data::Table*. This is mandatory, with no default value.

- **SIG_FILE**: directory and name of the file that contains the expected weak lensing signal (shear profile) from a halo. It is an ASCII file with HE_LI_SIG lines of header, then a table of data. Each row is a point where we evaluate the signal function, a column is the radial distance from zero at which the profile is evaluated and another is its values at that point. This is optional, to ignore set to "./" (default) or remove the parameter line.

- **COL_SIG_R**: number of the column of the radial position, in arcminutes. This is mandatory only if SIG_FILE is given, default is $-1$.

- **COL_SIG**: number of column of the signal value. This is mandatory only if SIG-FILE is given, default is $-1$.

- **HE_LI_SIG**: number of header lines to be skipped when reading the file with the *CosmoBolognalib* class *cbl::data::Table*. This is mandatory only if SIG_FILE is given, default is $-1$.

- **FKER_FILE**: directory and name of the file that contains the filter function in the Fourier domain. It is an ASCII file with FHE_LI_KER lines of header, then a table of data. Each row is a point where we evaluate the filter function; the column features are the mode $k$ at which the filter is evaluated and the value of the function in that point. This is mandatory, with no default value.

- **FCOL_KER_R**: number of the column of the mode $k$. This is mandatory, with no default value.

- **FCOL_KER**: number of column of the filter value. This is mandatory, with no default value.

- **FHE_LI_KER**: number of header lines to be skipped when reading the file with the *CosmoBolognalib* class *cbl::data::Table*. This is mandatory, with no default value.

# Appendix B

# Full analysis of catalogues C and D

We report here the full analysis of the results we obtained when running GtH21 on our data with the "customised" filter described in section 4.3.

## B.1 Application of the filter

When we run GtH21 we applied the same division into sub-fields discussed in section 4.2 and shown in figures 4.3c and 4.3d. We thus obtained nine output files. The E-mode and B-mode of the output map of the upper-central region are shown in figure B.1 as an example of a typical output. We then computed the strong and weak S/N thresholds above which a pixel is assumed to contain a detection as described in section 3.2.4 and as we show in figure B.2. We computed the histogram of the S/N distribution of the pixels of each sub-field with the same binning and we summed them to get the S/N distribution of all the pixels of simulation's field, for both the T-maps and the R-maps, that we plot in figure B.2a. The strong threshold is the maximum S/N of the R-map; for our data and the filter described in this appendix it is 7.036. In figure B.2b we plot the ratio between the counts of the T-map divided by the total number of counts (T-map plus R-map). This is used to identify the weak threshold, since we defined it as the S/N value above which the T-map detections are always more than 90% of the total (T-map plus R-map). For our data and the filter considered in this section it is 5.614.

Once the S/N thresholds were defined, we run the halo detection algorithm as described in section 3.2.4. Every group of contiguous pixels whose S/N is above the threshold is considered as a detection. The position of the detection is the pixel with the maximum S/N of the group, its S/N is the S/N of the detection. With this algorithm we have built two catalogues of detections; we included in a catalogue for each detection its position, its S/N, the number of pixels covered by it, its area on the sky in arcmin. As we discussed in section 4.3 the catalogue built with the strong threshold is called **catalogue C**, the one built with the weak threshold is called **catalogue D**.
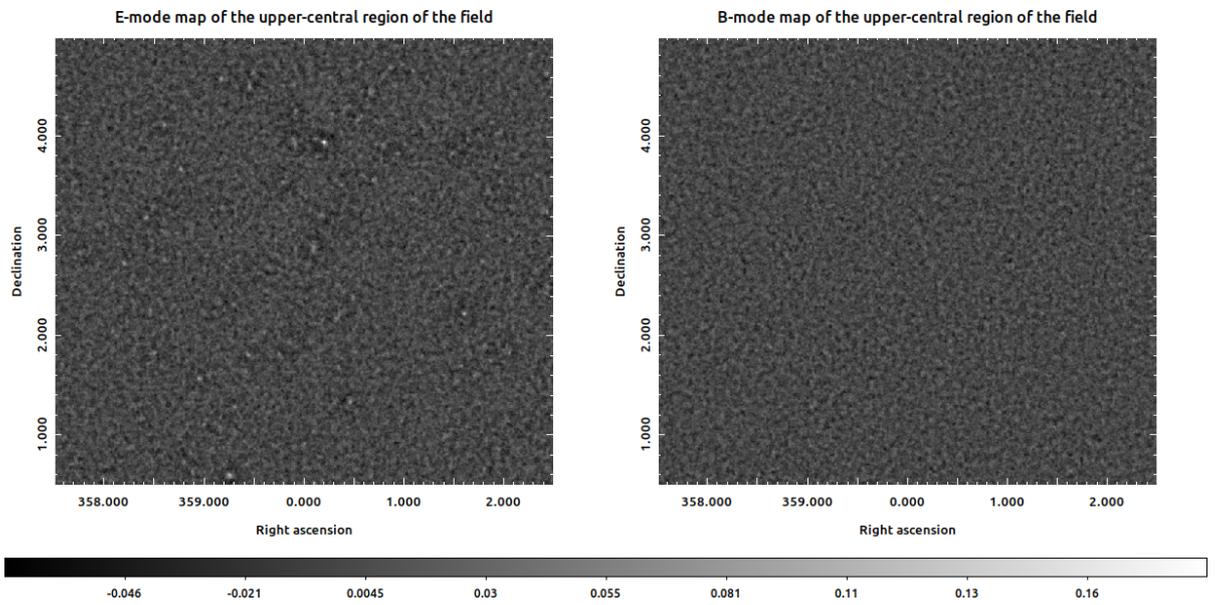
**Figure B.1:** Output maps of the upper-central region of the simulation, the E-mode is on the left, the B-mode on the right. They are plotted with the same color scale. In the E-mode map it is possible to see the typical signatures of a halo detection: high-values spots surrounded by lower-values troughs. These are not present in the B-mode, that only contains the noise from galaxies and the large scale structure.
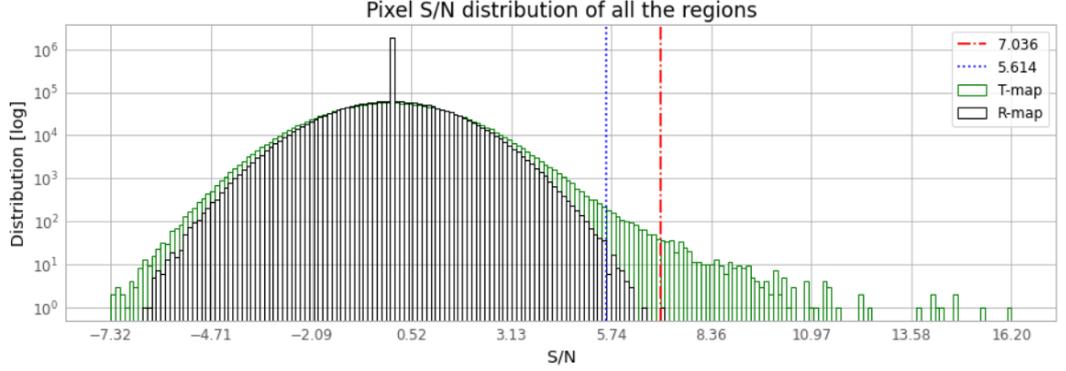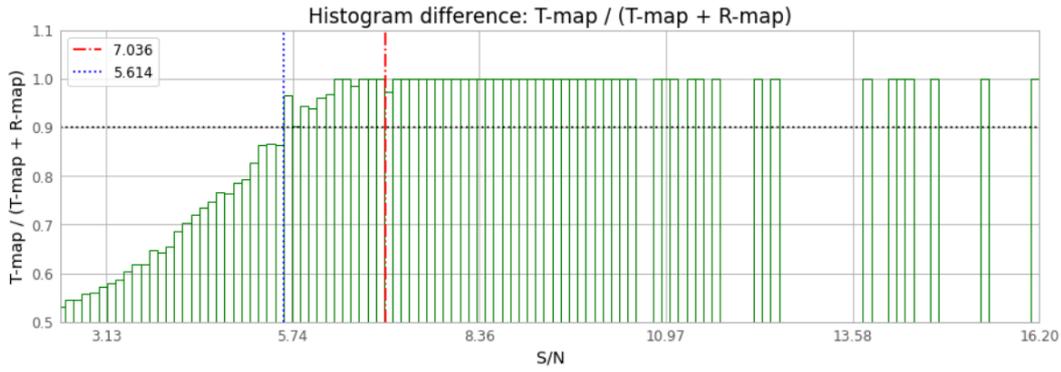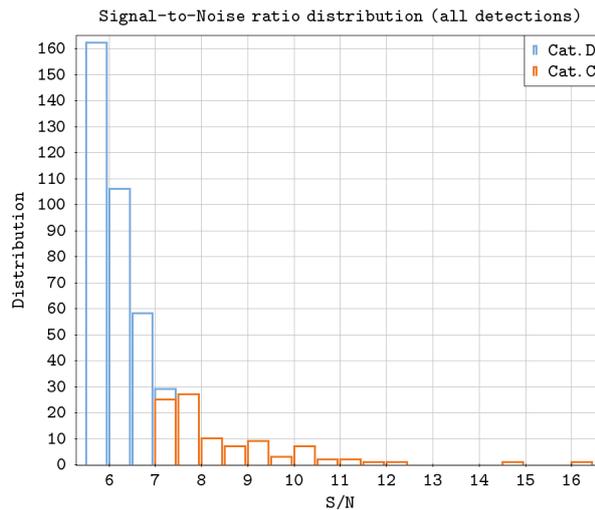
**(a)** S/N Distribution



**(b)** Threshold identifier

**Figure B.2:** Histograms used to define the strong and weak thresholds, as described in section 3.2.4. **Panel B.2a** shows the distribution of the S/N of all the pixels of the entire field of the simulation. The green histogram is the T-map (S/N of the E-mode, that uses the tangential shear), the black histogram is the R-map (S/N of the B-mode, that uses the cross shear). It is possible to see the high-S/N tail of the detections. For our data and the filter considered in section B.1 the strong threshold is 7.036; it is shown as a red line on both panels. In **panel B.2b** we use the same binning as the upper panel to plot, for each bin, the ratio between the counts of the T-map divided by the total number of counts (T-map plus R-map). This is used to identify the weak threshold, that is 5.614; it is shown as a blue line on both panels.

### B.1.1 Catalogues C and D

Catalogue $C$ is built with a S/N threshold of 7.036, it is made up by 96 detections. Catalogue $D$ is built with a S/N threshold of 5.614, it is made up by 426 detections. Catalogue $D$ can be thought as an extension of $C$ at lower S/N, as every detection in $C$ can also be found in $D$, with the same position and S/N. The ID numbers of the same detection in the two catalogues will in general be different, and the area of a detection in cat. $D$ will be larger than the area of the same detection in $C$, because we are allowing pixels with S/N between the two thresholds to be considered as part of a detection in $D$. In these catalogues there are no examples of "fragmented" detections by noise (detections present in $C$ but not in $D$) as in catalogues $A$ and $B$, that we discussed in section 4.2.

The S/N distribution of all the detections of catalogues $C$ and $D$ can be seen in figure B.3, where it becomes evident that catalogue $D$ is an extension of $C$. The spatial distribution of the detections of the two catalogues can be seen in figure B.4, where we also plot which detections have been matched to a simulated halo, as it is going to be described in the following section B.2.



**Figure B.3:** Signal-to-Noise ratio distribution of all the detections in catalogues $D$ (light blue histogram) and $C$ (orange histogram). Each bin is 0.5. Catalogue $C$ has been superimposed onto $D$; this shows graphically that from the S/N bin $[7.5; 8.0]$ the two catalogues have the same detections, as we expected.

## B.2 Matching catalogues C and D

We matched our detections in catalogues $C$ and $D$ with the simulated haloes by using *TOPCAT* Sky algorithm once again, with a maximum mismatch distance of $1.5'$. *TOPCAT*

matched 81 out of 96 detections (84.4%) of catalogue *C* and 221 out of 426 detections (51.9%) of catalogue *B*. The distribution of the detected and matched sources can be seen in figure B.4.

Figure B.5 shows the *detection-halo mismatch*, i.e. the distance vector between the matched detections and their halo. Figure B.5a shows the Right ascension and Declination components of the mismatch vectors, while figure B.5b shows the modulus of the vector, i.e. the distribution of the halo-detection distance. Figure B.5a tells us that GtH21, the detection and matching algorithms do not introduce systematic errors in the position of the detections. In fact, the matched detections of cat. *C* have average right ascension mismatch of $0.19' \pm 0.36'$ (error is the standard deviation) and average declination mismatch of $-0.14' \pm 0.66'$. The matched detections of cat. *D* have average right ascension mismatch of $0.18' \pm 0.38'$ and average declination mismatch of $-0.12' \pm 0.62'$. These average shifts are all consistent with zero, so they can be neglected. Figure B.5b show that the distributions of the mismatch distance are peaked and without tails, so $1.5'$ is a good choice for the maximum mismatch distance allowed. The average value ($\pm$ the standard deviation) of the mismatch distance distribution is $0.70' \pm 0.37'$ for catalogue *C* and $0.68' \pm 0.34'$ for catalogue *D*. The distributions we see in figure B.5b are peaked as expected.

All the detections matched in catalogue *C* have been matched in catalogue *D* with the same halo of the simulation, except for one detection: source *C#*23. This source is matched with halo *H#*861. In catalogue *D* source *C#*23 is detected and registered as source *D#*100, where it is *not* matched to any halo at all. This is because in catalogue *D* one of the detections added because of the lower threshold is *D#*99, that is the one matched with *H#*861. *TOPCAT*'s matching algorithm does not take into account the S/N of a detection in the matching process, but only the position of the haloes and the detections. Halo #861 is then matched to detections with different positions in cat. *C* and *D* because it is closer to detection *D#*99 (distance 0.52 arcmin) than *D#*100 (distance 0.89 arcmin), despite the lower S/N. These sources are described in table B.1, where we see that *D#*99 has a lower S/N than *D#*100 (5.66 and 9.32 respectively) and it is made up by a single pixel. These values lead us to believe that *D#*99 is either a weaker part of detection *D#*100 that has been fragmented due to noise and our simple algorithm to detect haloes, or a single-pixel spurious peak that happens to arise between *D#*100 and *H#*861. A more sophisticated matching algorithm that takes S/N and source extension into account would probably match *H#*861 with *D#*100. Since this is the only "mismatched" source of the catalogues and it arises because of our simple detection and matching algorithms we shall not consider it to be statistically relevant for the rest of the chapter.
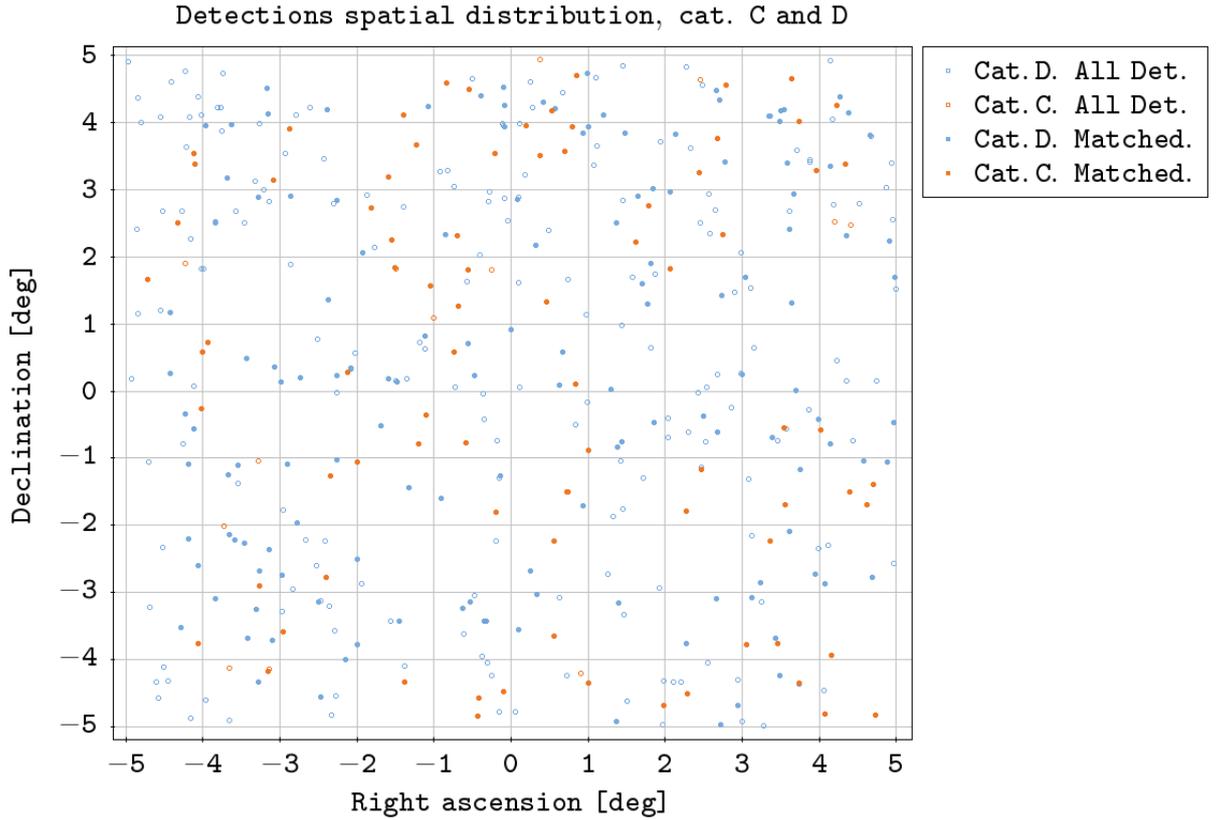
**Figure B.4:** Spatial distribution of the detections. Orange circles are the detections of catalogue *C* (S/N threshold 7.036). When the circle is empty the detection has not been matched, when the circle is full it has been matched with a halo. These sources are also included in catalogue *D* (S/N threshold 5.614). Light blue circles are the detections of catalogue *D* only, not found in catalogue *C*. When the circle is empty the detection has not been matched, when the circle is full it has been matched with a halo.

**(a)** Spatial distribution.

**(b)** Distance distribution.
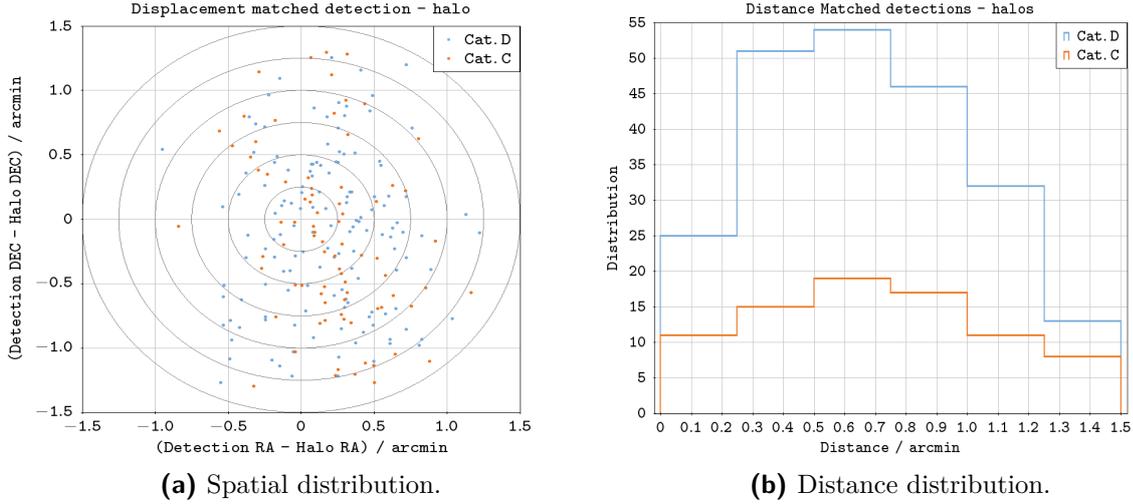
**Figure B.5:** Distribution of the detection-halo mismatch. In **panel B.5a** we show the distribution of the right ascension and declination components of the mismatch vector, defined as the position of the detection minus the position of its matched halo. Orange circles are the detections found in both catalogues $C$ and $D$, light blue circles only in $D$. We also plot circles of fixed radius around $(0, 0)$ increasing the radius by 0.25: the counts in each annulus are the height of the bins shown in **panel B.5b**. There we show the distribution of the *modulus* of the mismatch, i.e. the detection-halo distance. The orange step-histogram is relative to cat. $C$, the light blue step-histogram is for cat. $D$. Their maximum value is $1.5'$, the limit set for the matching algorithm. We expected these distribution to be approximately Gaussian. The average and standard deviation of the distributions are $0.70' \pm 0.37'$ for cat. $C$ and $0.68' \pm 0.34'$ for cat. $D$. We obtain a peaked distribution as expected.

**Table B.1:** Detections $C\#23$, $D\#99$, $D\#100$ and halo $H\#861$. We report their main properties. Halo $H\#861$ has been matched to $C\#23$ and $D\#99$, despite the equivalent detection of $C\#23$ in catalogue $D$ is $D\#100$. This is because *TOPCAT*'s algorithm take only the position of the sources into account to match them. We believe that $D\#99$ is either a weaker part of detection $D\#100$ that has been fragmented due to noise and our simple algorithm to detect haloes, or a single-pixel spurious peak that happens to arise between $D\#100$ and $H\#861$. See section B.1.1 for a more accurate discussion.

**(a)** Halo $H\#861$.

| ID | RA [deg] | DEC [deg] | Mass [$10^{14}\,\mathrm{M_\odot}\,h^{-1}$] | Redshift |
|----|----------|-----------|-------------------------------------------|----------|
| 861 | 3.74678 | -4.35485 | 4.14 | 0.343 |

**(b)** Detections $C\#23$, $D\#99$, $D\#100$.

| ID | RA [deg] | DEC [deg] | S/N | Area [pxl] |
|----|----------|-----------|-----|------------|
| $C\#23$ | 3.74025 | -4.34152 | 9.3179 | 5 |
| $D\#100$ | 3.74025 | -4.34152 | 9.3179 | 9 |
| $D\#99$ | 3.74753 | -4.36344 | 5.6630 | 1 |

## B.2.1 Signal-to-noise ratio

The distribution of the S/N can be seen in figure B.6 for both catalogues $C$ and $D$. They show that most of the unmatched detections are introduced when we lower the detection threshold from 7.036 to 5.614. To be more precise, 189 out of the 205 unmatched detections in catalogue $D$ have S/N $< 7.036$. As we found in section 4.2.2, confirming the results of Pace et al. (2007), a S/N threshold $\simeq 7$ can once again be considered as a reliable threshold to detect haloes trough weak lensing: 80 out of the 96 detections of catalogue $D$ with S/N $> 7$ are matched. These histograms show graphically that the two catalogues have the same detections and matches starting from the S/N bin $[7.5; 8.0]$ as expected, apart from one matched detection missing from cat.$D$ in bin $[9.0; 9.5]$. This is of course $D\#100$, that has not been matched to $H\#861$ as its counterpart $C\#23$ because *TOPCAT* assigned that halo to $D\#99$, a detection with S/N 5.66. We believe that more sophisticated halo detection and matching algorithms would probably detect only one source centered in the same position of $D\#100$ and match it correctly with $H\#861$, just as it is matched with $C\#23$ in catalogue $C$.

We study once again the *purity* of our catalogues. In figure B.7 we show the percentage of matched detections with S/N above a given value. In order to plot them we computed the total number of detections and number of matched detections *above* a given S/N in each of our catalogues and divided them. Above S/N= 7.0 the percentage of matched

detections is always above 85% for catalogue $C$, so this can be considered as a reliable S/N value for the construction of a catalogue; this is consistent with the results of Pace et al. (2007). The curve of catalogue $D$ is lower because of source $D\#100$, so we consider the curve of catalogue $C$ to better represent our results after S/N= 7.5. A purity of 100% is reached for sources with S/N $> 10.5$.

The unmatched sources with high S/N are probably spurious detections caused by the large scale structure. Roughly speaking, $\approx 17\%$ of the detected haloes with S/N $> 7$ might be spurious peaks caused by weak lensing from the large scale structure.

In figure B.8 we show the S/N distribution of the matched detections as a function of mass and redshift for catalogues $C$ (figure B.8a) and $D$ (figure B.8b). They show that we find detections with higher S/N in bins of redshift $\in [0.2; 0.4]$ and mass $> 4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$. This suggests that our linear filter is optimized to detect haloes with these redshifts and masses, despite the fact that we set the lens redshift parameter of the filter to be 0.6. The filter might not properly detect haloes at redshift 0.6 because their lensing signal is not as strong as the one of equally massive haloes at redshift 0.3. The filter detects haloes more efficiently if these are at redshifts between 0.2 and 0.5, i.e. at intermediate distances between the observer and the sources (Pace et al. 2007).

**(a)** Catalogue *C*.

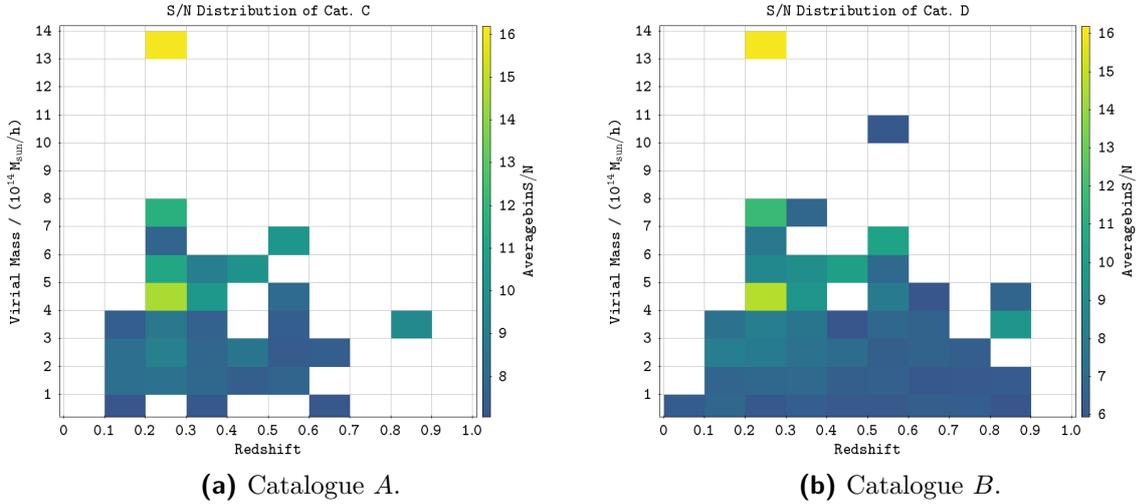**(b)** Catalogue *D*.
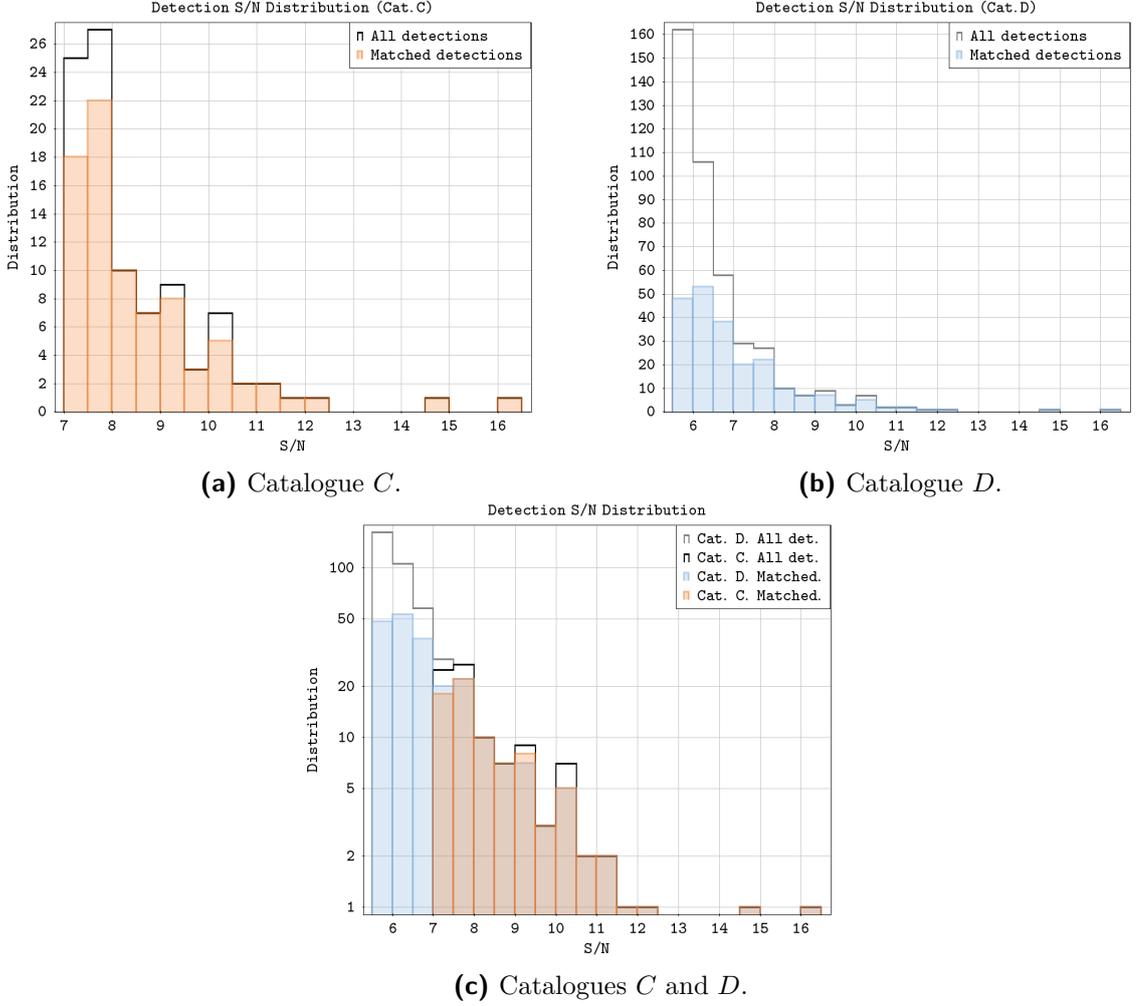
**(c)** Catalogues *C* and *D*.

**Figure B.6:** Distribution of the signal to noise ratio for all the detections and the matched detections of the two catalogues. Catalogue *C* is shown in **panel B.6a**, catalogue *D* is shown in **panel B.6b**. They are both shown in **panel B.6c** for comparison. These histograms show graphically that the two catalogues have the same detections and matches starting from the S/N bin $[7.5; 8.0]$ as expected, apart from one matched detection missing from cat. *D* in bin $[9.0; 9.5]$. We do not consider this to be statistically relevant for this work, see section B.2.
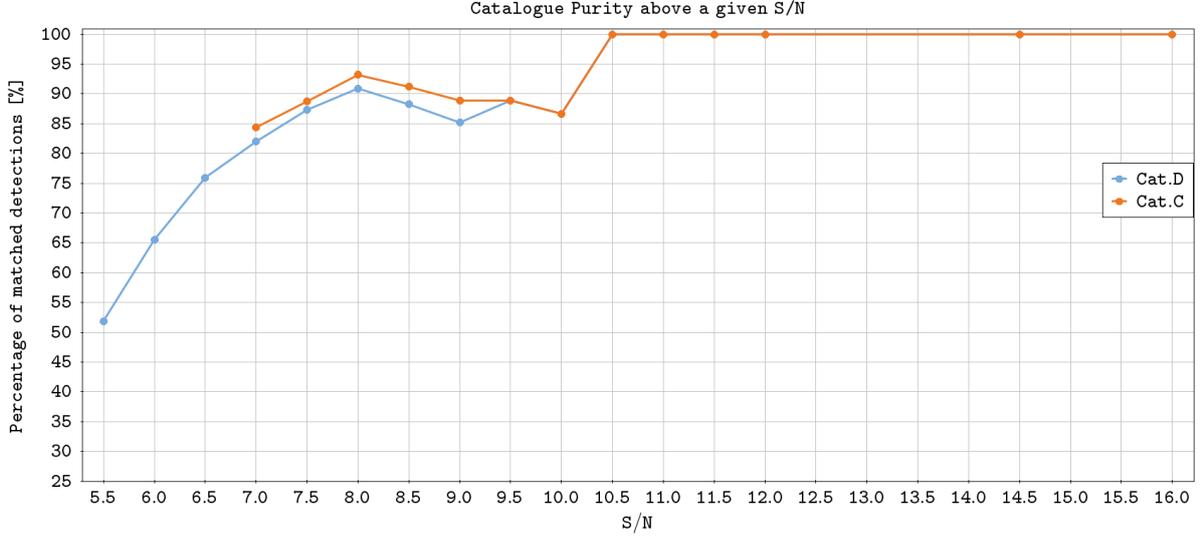
**Figure B.7:** Percentage of matched detections above a given S/N. For instance, 92% of the detections with S/N > 8.50 are matched in catalogue *C*. The light blue curve of cat. *D* has slightly lower values than the orange curve of cat. *C* until S/N 9.5 because of source *D#100*. We thus believe the curve of catalogue *C* is the most accurate representation of our results.



**(a)** Catalogue *A*.

**(b)** Catalogue *B*.

**Figure B.8:** S/N ratio distribution of the matched detections as a function of mass and redshift. We plot the average S/N of the matched detections of catalogue *C* (**panel B.8a**) and *D* (**panel B.8b**) in each bidimensional bin (redshift bins $\Delta z = 0.1$, mass bins $\Delta M = 10^{14}\,\mathrm{M}_\odot\,h^{-1}$). In both catalogues the average S/N is higher when the redshift is $\in [0.2; 0.5]$ and the mass is $> 4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$.

## B.2.2 Redshift

The redshift distribution of the matched haloes with respect to all the simulated haloes is shown in figure B.9, while figure B.10 shows the percentage of matched haloes for each redshift bin of figure B.9. The median redshift of the matched detections is 0.33 with quartiles $0.25 \div 0.40$ for catalogue $C$ and 0.35 with quartiles $0.26 \div 0.49$ for catalogue $D$, while the median redshift of all the simulated haloes is 0.62 with quartiles $0.46 \div 78$. The filter selects haloes at intermediate distances between the observer and the sources, because of the dependence of the geometrical lensing strength on the angular-diameter distances between the observer and the lens, the lens and the sources, and the observer and the sources (Pace et al. 2007). The completeness of our catalogues, i.e. the percentage of matched haloes, has a peak extended from redshift $\approx 0.15$ to $\approx 0.35$, as it is shown in figure B.10. Another observation from this figure is that completeness systematically increases in catalogue $D$; as we already found in section 4.2.2 for catalogues $A$ and $B$ lowering the threshold introduces a relevant fraction of reliable detections despite the matching fraction of the overall catalogue decreases.

We expect the completeness to be related to the mass: at a given redshift, more massive lenses are expected to produce stronger lensing signal. We plot in figure B.11 the catalogue completeness as a function of redshift, with different cuts in mass. In particular in figure B.11a we have plotted the completeness of catalogue $C$ as a function of redshift for 4 subset of haloes: haloes with virial mass $< 1 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$, haloes with mass $\in [1; 2] \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$, haloes with mass $\in [2; 4] \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$, haloes with mass $> 4 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$. We do the same in figure B.11b for catalogue $D$. Figure B.11c shows the completeness of catalogue $C$ as a function of redshift for the haloes with 4 different cuts in mass: haloes with mass $> 4 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$, haloes with mass $> 2 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$, haloes with mass $> 1 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$, all the haloes. We do the same in figure B.11d for catalogue $D$. These plots also confirm the existence of a peak extended around redshift $\approx 0.15 \div 0.35$ and that the completeness of catalogue $D$ is higher than catalogue $C$, hence using the weak threshold can actually improve the completeness of our catalogue. They also show that, as a general trend, when the mass increases the matching percentage increases too, although this is not always the case as there are some exceptions. This result is consistent with what we found in section 4.2.2. The massive haloes with mass $> 4 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$ show significant fluctuations due to their small number: especially at $z < 0.2$ and $z > 0.6$ not all of them are detected. Let us remember that these haloes have poor statistics: there are 31 haloes in the simulation with mass $> 4 \times 10^{14} \, \mathrm{M}_\odot \, h^{-1}$ and we match 15 in cat. $C$ and 21 in $D$ out of them all. A high background noise fluctuation that lowers the S/N might be responsible not detecting some massive haloes, with a resulting strong loss of percentage. As we remarked in section 4.2.2 losing massive haloes at low and high redshift might also be a suggestion that some of the lower mass matched detections may be random matches between a halo and a spurious detection caused by noise. Figure B.11 also shows that the peak of completeness around redshift $\approx 0.15 \div 0.35$

is always found.

We plot the distribution of the matched haloes with respect to all the haloes in figure B.12, where it becomes clear that most of the matched detections have intermediate redshift $\approx 0.2 \div 0.5$ and intermediate-high mass $> 2 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$. The distribution in figure B.12 is turned into maps of completeness in figure B.13. In order to realize them we counted, for each bidimensional bin, the number of matched haloes in catalogues $C$ (figure B.13a) and $D$ (figure B.13b) and divided by the total number of simulated haloes. The haloes have been grouped into bins of mass $\Delta M = 1 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$ and redshift $\Delta z = 0.1$. These plots show that completeness increases in the region of intermediate redshift $\approx 0.2 \div 0.5$ and intermediate-high mass $> 3 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$, a result consistent with what we found in section 4.2.2.



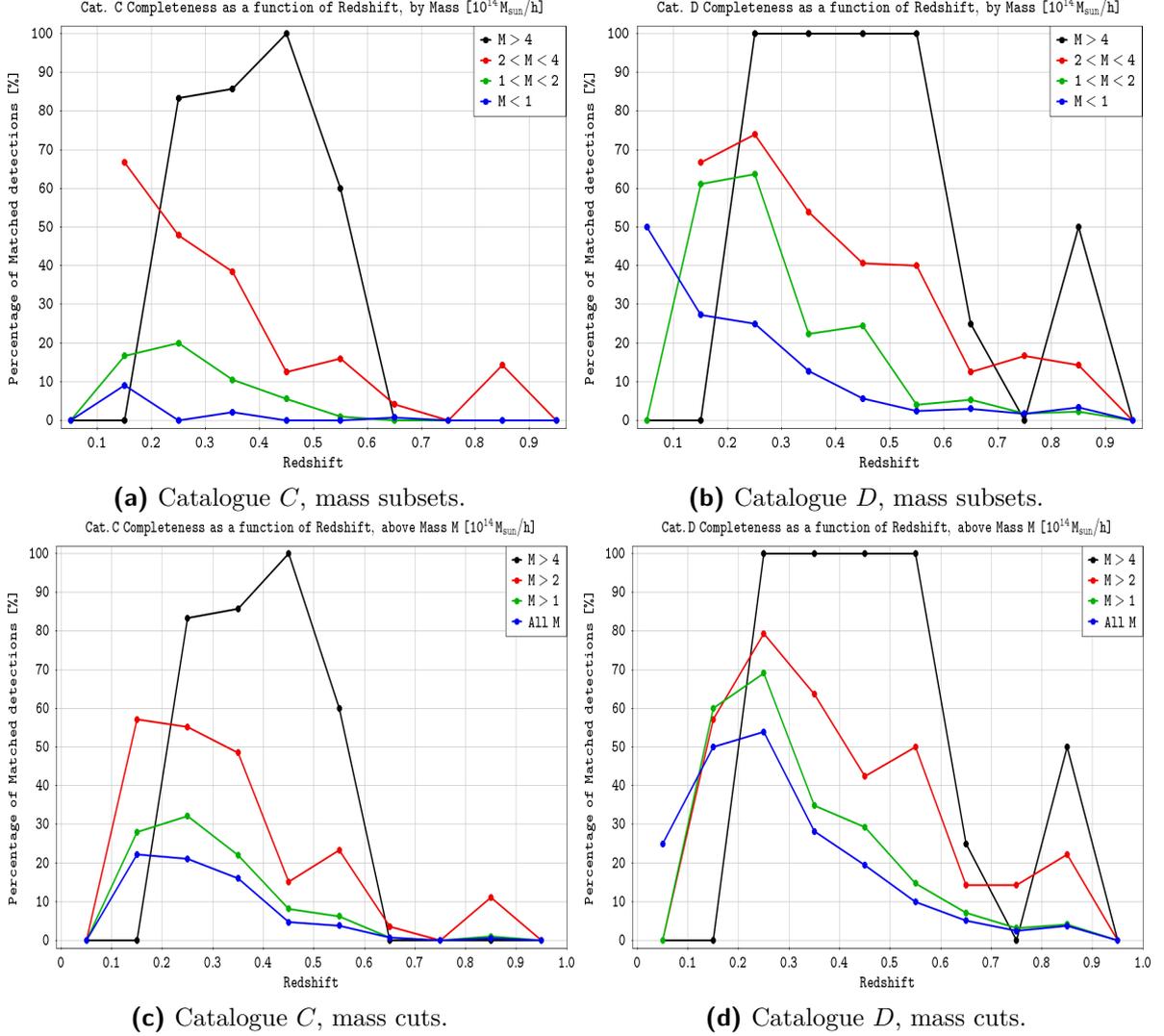**(a)** Catalogue $C$          **(b)** Catalogue $D$
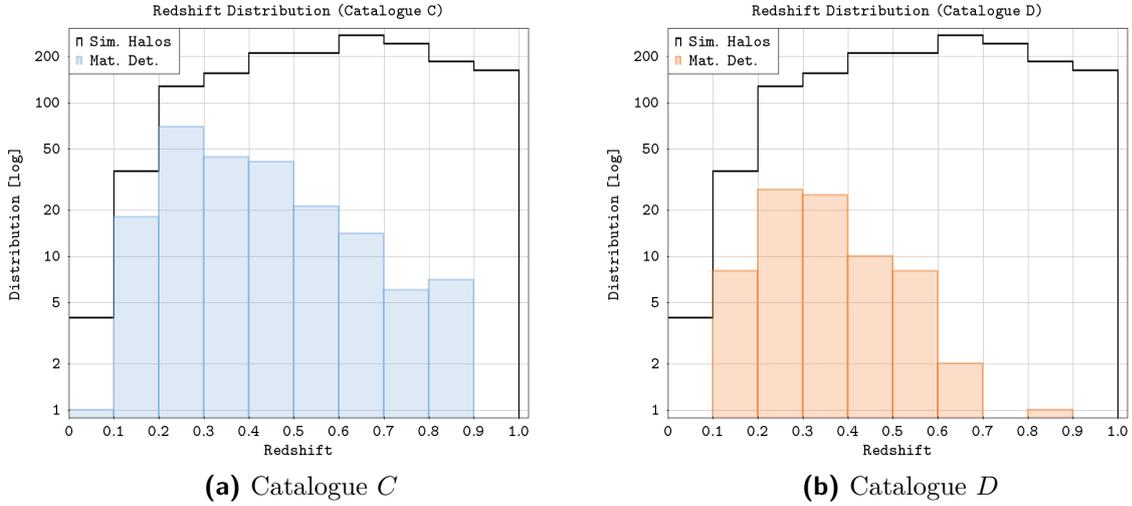
**Figure B.9:** Redshift distribution of the matched detections and total simulated haloes.
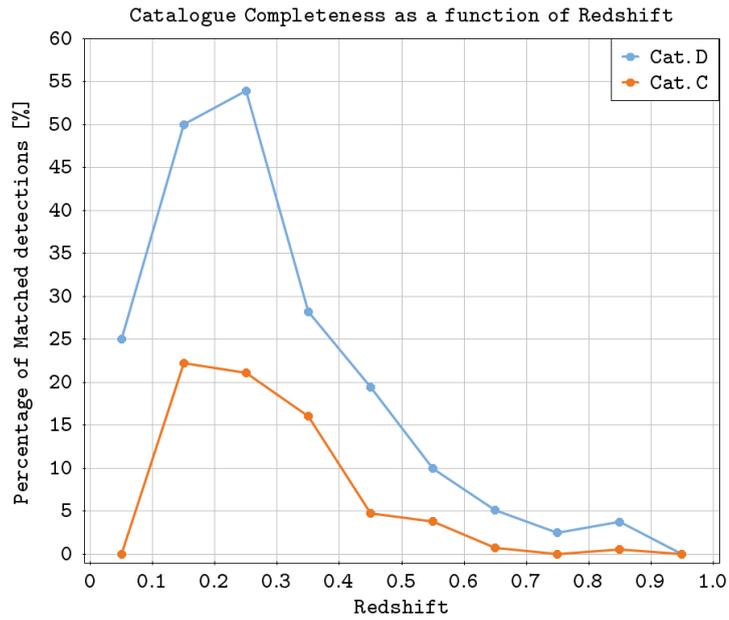
**Figure B.10:** Catalogue completeness (percentage of matched haloes) as a function of redshift in catalogue *C* (orange line) and *D* (light blue line). This includes all the haloes with no mass cut applied. To get these points we divided the number of matched haloes by the total number of haloes in each redshift bin used in figure B.9, and we placed the points at the center of their bin. There is a clear peak extended between redshift $\approx 0.15$ and $\approx 0.35$ in both the catalogues: the filter does not detect the haloes with the same efficiency at any redshift. It selects the haloes with intermediate distance between observer and nearby background sources. The matching percentage increases in catalogue *D*; this suggests that lowering the threshold introduces a relevant fraction of reliable detections.

**(a)** Catalogue *C*, mass subsets.

**(b)** Catalogue *D*, mass subsets.

**(c)** Catalogue *C*, mass cuts.

**(d)** Catalogue *D*, mass cuts.

**Figure B.11:** Catalogue completeness (percentage of matched haloes) as a function of redshift, by virial mass. In **Panela B.11a** (cat. *C*) and **B.11b** (cat. *D*) we plot the catalogue completeness as a function of redshift for 4 subsets of the halo catalogue according to their mass. In **Panels B.11c** (cat. *C*) and **B.11d** (cat. *D*) we plot the catalogue completeness as a function of redshift for 4 subsets of the halo catalogue with different cuts of mass. As a general trend the four plots show that completeness increases with mass and it is higher in cat. *D* than *C*. The main exceptions are the massive haloes $> 4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$, that show significant fluctuations: at $z < 0.2$ and $z > 0.6$ not every halo is detected and we can detect lower percentages than the lower mass haloes. This might be due to their poor statistics, or it might suggests that some of the lower mass matched detections are actually random matches between a halo and a spurious detection caused by noise. These plots show the peak extended around redshift $\approx 0.15 \div 0.35$: the efficiency of detecting haloes has a strong dependence on redshift.

117
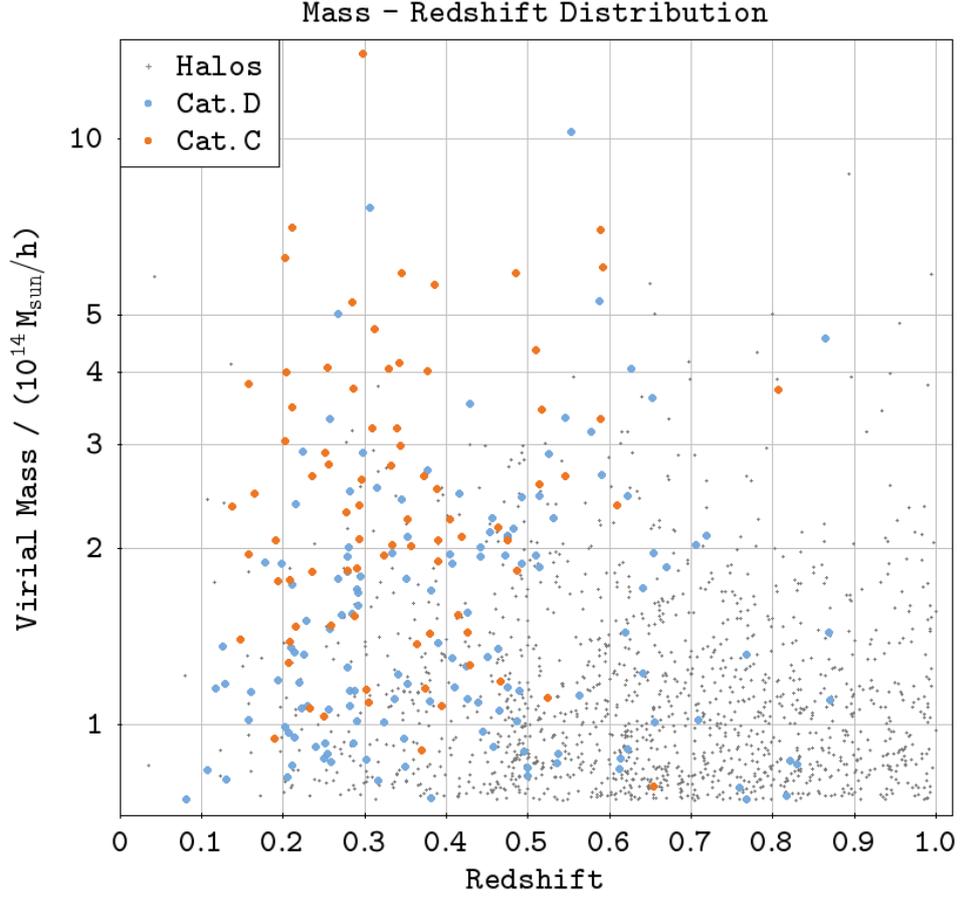
**Figure B.12:** Distribution of the haloes as a function of Mass and Redshift. Grey crosses are all the simulated haloes. Light blue circles are the matched detections of catalogue $D$, superimposed to the cross of their halo. Orange circles are the matched detections of catalogue $C$, superimposed to their counterparts in catalogue $D$ and halo. Most of the matched haloes have intermediate redshift $\approx 0.2 \div 0.5$ and intermediate-high mass $> 2 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$.

**(a)** Catalogue $C$.  **(b)** Catalogue $D$.
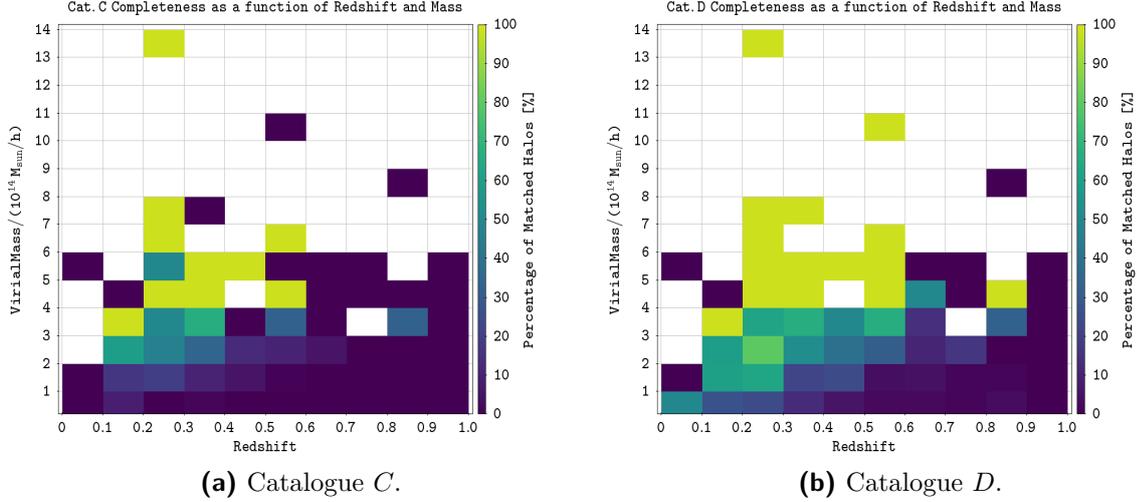
**Figure B.13:** Percentage of matched haloes (completeness) as a function of mass and redshift for the two catalogues. Each bidimensional bin shows the percentage of matched detections (over all the simulated haloes) in that bin. The percentages are shown as a colour map for catalogue $C$ and $D$ in **panel B.13a** and **panel B.13b** respectively. Most of the haloes with intermediate redshift $\approx 0.2 \div 0.5$ and intermediate-high mass $> 3 \times 10^{14}\,\mathrm{M_\odot}\,h^{-1}$ are matched.

### B.2.3 Virial Mass

The virial mass distribution of the matched haloes is shown in figure B.14, where it is compared with the distribution of all the simulated haloes. Not every mass can be detected with the same efficiency: in units of $10^{14}\,\mathrm{M_\odot}\,h^{-1}$ the median mass of the simulated haloes is 1.07 with quartiles $0.86 \div 1.47$, while the median mass of the matched detections is 2.45 with quartiles $1.53 \div 3.44$ in catalogue $C$ and 1.76 with quartiles $1.10 \div 2.46$ in catalogue $D$. As it happened for catalogues $A$ and $B$ in section 4.2.2 the second most massive haloes of the simulation (halo $H\#12$) is detected and matched only in catalogue $D$; the third most massive (halo $H\#1$) in not detected in either of them. They are described in table 4.2 of chapter 4. A notable feature in catalogue $C$ is that the fourth massive halo of the simulation is not matched: this is halo $H\#194$, with a mass of $8.26 \times 10^{14}\,\mathrm{M_\odot}\,h^{-1}$ at redshift 0.31. It was matched in catalogue $A$ with source $A\#97$ that had S/N$= 6.976$, so losing it can be attributed to an increase of threshold from 6.871 for cat. $A$ to 7.036 for cat. $C$. In fact, it is matched to detection $D\#338$ (and $B\#355$).

Figure B.15 describes catalogues $C$ and $D$ completeness as a function of the virial mass. Figure B.15a plots the percentage of matched haloes in the mass bins of figure B.14 at the lowest end of the bin. Figure B.15b shows the the percentage of matched detections with virial mass above a given value. At $4 \times 10^{14}\,\mathrm{M_\odot}\,h^{-1}$ we reach 65% of completeness for catalogue $D$. The percentages are steadily growing functions of the mass until we get to $4 \times 10^{14}\,\mathrm{M_\odot}\,h^{-1}$; after that the poor statistics makes them subject to
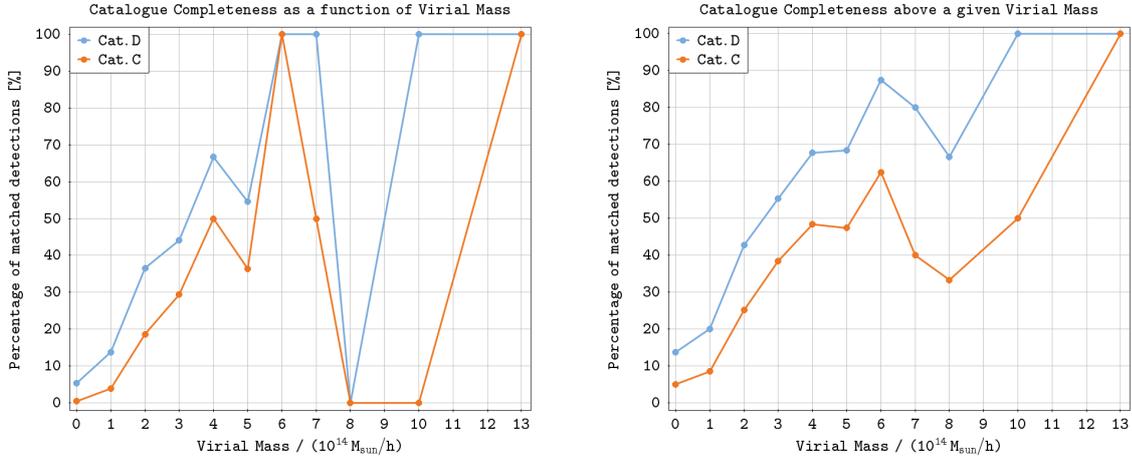
119

strong fluctuations. Those are responsible for the dips in completeness that can be seen at high mass values. There are in fact only 5 haloes in total with mass $> 7 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$; catalogue $C$ does not detect three of them ($H\#12$, $H\#1$, $H\#194$) while $D$ does not detect $H\#1$. This prevents the completeness functions from reaching 100% but in the last bin (due to the detection of the most massive halo of the simulation in both catalogues). The completeness is lower for catalogue $C$, another confirmation that lowering the threshold can in fact improve the catalogue completeness.

The catalogue completeness is a function of both the virial mass and redshift; we expect the completeness to be higher at intermediate redshift. We have plotted in figure B.16 the completeness of our catalogues as a function of mass or above a given mass after we had divided the haloes in four subsets according to their redshift: $z < 0.2$, $0.2 < z < 0.4$, $0.4 < z < 0.6$, $z > 0.6$. We plot the completeness as a function of mass in figures B.16a (catalogue $C$) and B.16b (catalogue $D$); we have used the same binning of figure B.15a. We plot completeness above a given virial mass in figures B.16c (cat. $C$) and B.16d (cat. $D$), splitting each curve of figure B.15b into three redshift subsets. For $z < 0.2$ and $z > 0.6$ our filter does not find many haloes, hence the completeness above a given mass never goes above 55%. For haloes with $0.2 < z < 0.6$ the matching percentage above a given mass increases when the mass increases and it reaches 100% completeness at mass $4 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$ in cat. $D$. It is not reached in cat. $C$ until the very last bin. As we found in section 4.2.2 the haloes with $0.2 < z < 0.6$ have the greatest probability of being detected.
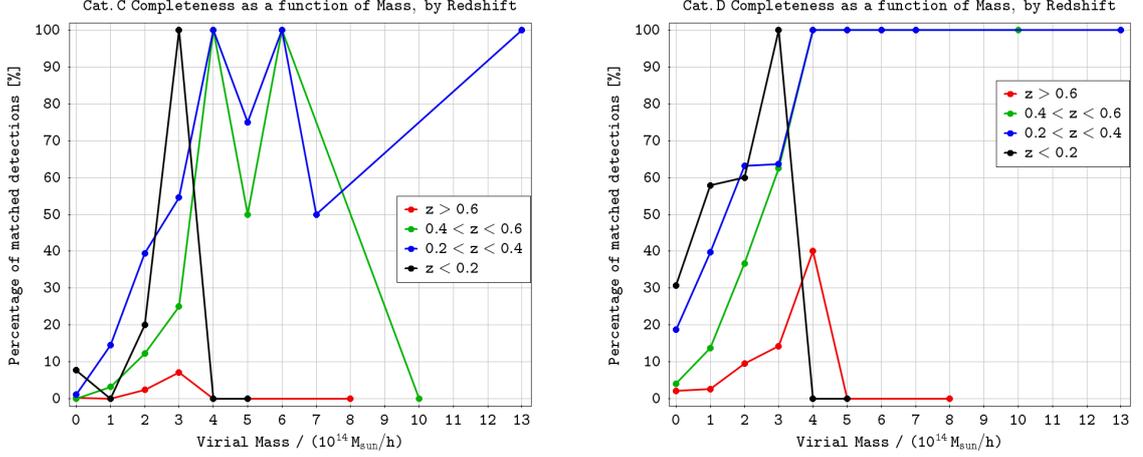
**(a)** Catalogue *C*
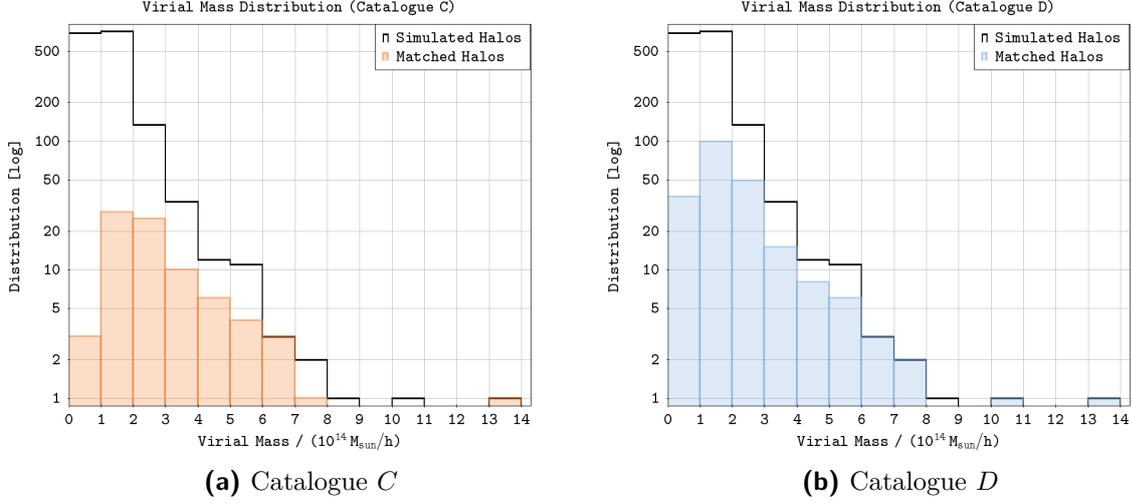
**(b)** Catalogue *D*

**Figure B.14:** Distribution of the virial mass $M_{200}$ for the matched haloes of catalogue *C* (**panel B.14a**), *D* (**panel B.14b**), and all the simulated haloes. Note that the second ($H\#12$) and fourth ($H\#194$) most massive halo are detected and matched only in catalogue *D*, while the third ($H\#1$) is lost in both catalogues. This can be probably attributed to their high redshift for $H\#12$ ($z = 0.55$) and $H\#1$ ($z = 0.89$) and the more restrictive S/N threshold of catalogue *D* for $H\#194$.
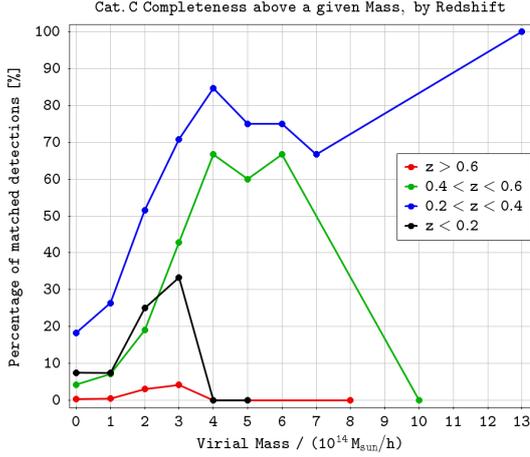


**(a)** Percentage of matched haloes as a function of the virial mass.

**(b)** Percentage of matched haloes above a given virial mass.

**Figure B.15:** Percentage of matched haloes as a function of the virial mass (**panel B.15a**) and above a given virial mass (**panel B.15b**). Both graphs show that the catalogue completeness increases when the mass increases. The percentages at higher masses than $> 7 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$ are computed on a very small number of haloes, hence they are subject to strong fluctuations and a dip in percentages due to losing some of the most massive haloes of the simulation.
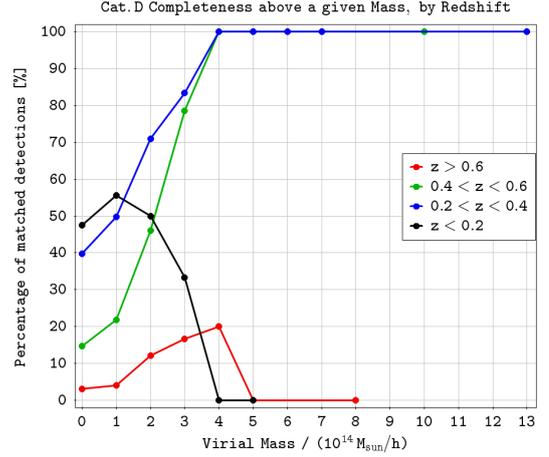
**(a)** Percentage of matched haloes as a function of the virial mass, by redshift, catalogue $C$.

**(b)** Percentage of matched haloes as a function of the virial mass, by redshift, catalogue $D$.

**(c)** Percentage of matched haloes above a given virial mass, by redshift, catalogue $C$.

**(d)** Percentage of matched haloes above a given virial mass, by redshift, catalogue $D$.

**Figure B.16:** Percentage of matched haloes as a function of the virial mass, by redshift. These graphs show how the completeness of the catalogues changes as a function of halo redshift. The sources with redshift $z \in [0.2; 0.6]$ tend to have the highest matching percentages because of the peak in detection efficiency of the filter in this redshift range. Another trend is that the matching percentage tends to increase as a function of the halo mass, although there are fluctuations given by the poor statistics.

# References

Aghanim, N. et al. (Sept. 2020). 'Planck 2018 results'. In: *Astronomy and Astrophysics* 641, A6. ISSN: 1432-0746. URL: http://dx.doi.org/10.1051/0004-6361/201833910.

Albrecht, Andreas et al. (Sept. 2006). 'Report of the Dark Energy Task Force'. In: *arXiv e-prints*, astro-ph/0609591, astro–ph/0609591. arXiv: astro-ph/0609591 [astro-ph].

Bartelmann, M. and P. Schneider (Jan. 2001). 'Weak gravitational lensing'. In: *Physics Reports* 340.4-5, pp. 291–472. arXiv: astro-ph/9912508 [astro-ph].

Bellagamba, Fabio et al. (Feb. 2018). 'AMICO: optimized detection of galaxy clusters in photometric surveys'. In: *Monthly Notices of the Royal Astronomical Society* 473.4, pp. 5221–5236. arXiv: 1705.03029 [astro-ph.CO].

Carroll, Sean M. (2004). *Spacetime and geometry. An introduction to general relativity.*

Chang, C. et al. (Sept. 2013). 'The effective number density of galaxies for weak lensing measurements in the LSST project'. In: *Monthly Notices of the Royal Astronomical Society* 434.3, pp. 2121–2135. arXiv: 1305.0793 [astro-ph.CO].

Crittenden, Robert G. et al. (Mar. 2002). 'Discriminating Weak Lensing from Intrinsic Spin Correlations Using the Curl-Gradient Decomposition'. In: *Astrophysical Journal* 568.1, pp. 20–27. arXiv: astro-ph/0012336 [astro-ph].

Erben, T. et al. (Oct. 2003). 'A weak lensing analysis of a STIS dark-lens candidate'. In: *Astronomy and Astrophysics* 410, pp. 45–52. arXiv: astro-ph/0307381 [astro-ph].

Ettori, S. et al. (Jan. 2006). 'The baryon fraction in hydrodynamical simulations of galaxy clusters'. In: *Monthly Notices of the Royal Astronomical Society* 365.3, pp. 1021–1030. arXiv: astro-ph/0509024 [astro-ph].

Euclid Collaboration et al. (July 2019). 'Euclid preparation. III. Galaxy cluster detection in the wide photometric survey, performance and algorithm selection'. In: *Astronomy and Astrophysics* 627, A23, A23. arXiv: 1906.04707 [astro-ph.CO].

Giocoli, Carlo, Marco Baldi and Lauro Moscardini (Dec. 2018). 'Weak lensing light-cones in modified gravity simulations with and without massive neutrinos'. In: *Monthly Notices of the Royal Astronomical Society* 481.2, pp. 2813–2828. arXiv: 1806.04681 (astro-ph.CO).

Giocoli, Carlo et al. (Sept. 2015). 'Disentangling dark sector models using weak lensing statistics'. In: *Monthly Notices of the Royal Astronomical Society* 452.3, pp. 2757–2772. arXiv: 1502.03442 [astro-ph.CO].

Hilbert, Stefan et al. (Mar. 2020). 'The accuracy of weak lensing simulations'. In: *Monthly Notices of the Royal Astronomical Society* 493.1, pp. 305–319.
arXiv: 1910.10625 (astro-ph.CO).

Kaiser, Nick and Gordon Squires (Feb. 1993). 'Mapping the Dark Matter with Weak Gravitational Lensing'. In: *Astrophysical Journal* 404, p. 441.

Marulli, F., A. Veropalumbo and M. Moresco (Jan. 2016). 'CosmoBolognaLib: C++ libraries for cosmological calculations'. In: *Astronomy and Computing* 14, pp. 35–42.
arXiv: 1511.00012 [astro-ph.CO].

Maturi, M. et al. (Nov. 2005). 'An optimal filter for the detection of galaxy clusters through weak lensing'. In: *Astronomy and Astrophysics* 442.3, pp. 851–860. arXiv: astro-ph/0412604 [astro-ph].

Maturi, M. et al. (Feb. 2007). 'Searching dark-matter halos in the GaBoDS survey'. In: *Astronomy and Astrophysics* 462.2, pp. 473–479.
arXiv: astro-ph/0607254 (astro-ph).

Meneghetti, Massimo, Matthias Bartelmann and Lauro Moscardini (Mar. 2003). 'Cluster cross-sections for strong lensing: analytic and numerical lens models'. In: *Monthly Notices of the Royal Astronomical Society* 340.1, pp. 105–114.
arXiv: astro-ph/0201501 (astro-ph).

Metcalf, R. Benton and Margarita Petkova (Dec. 2014). 'GLAMER - I. A code for gravitational lensing simulations with adaptive mesh refinement'. In: *Monthly Notices of the Royal Astronomical Society* 445.2, pp. 1942–1953. arXiv: 1312.1128 [astro-ph.CO].

Navarro, J. F. et al. (Apr. 2004). 'The inner structure of ΛCDM haloes - III. Universality and asymptotic slopes'. In: *Monthly Notices of the Royal Astronomical Society* 349.3, pp. 1039–1051.
arXiv: astro-ph/0311231 (astro-ph).

Navarro, Julio F., Carlos S. Frenk and Simon D. M. White (May 1996). 'The Structure of Cold Dark Matter Halos'. In: *Astrophysical Journal* 462, p. 563. arXiv: astro-ph/9508025 [astro-ph].

Pace, F. et al. (Sept. 2007). 'Testing the reliability of weak lensing cluster detections'. In: *Astronomy and Astrophysics* 471.3, pp. 731–742.
arXiv: astro-ph/0702031 (astro-ph).

Petkova, Margarita, R. Benton Metcalf and Carlo Giocoli (Dec. 2014). 'GLAMER - II. Multiple-plane gravitational lensing'. In: *Monthly Notices of the Royal Astronomical Society* 445.2, pp. 1954–1966. arXiv: 1312.1536 [astro-ph.CO].

Roncarelli, M. et al. (July 2007). 'The Sunyaev-Zel'dovich effects from a cosmological hydrodynamical simulation: large-scale properties and correlation with the soft X-ray signal'. In: *Monthly Notices of the Royal Astronomical Society* 378.4, pp. 1259–1269.
arXiv: astro-ph/0701680 [astro-ph].

Sachs, R. K. and A. M. Wolfe (Jan. 1967). 'Perturbations of a Cosmological Model and Angular Variations of the Microwave Background'. In: *Liege International Astrophysical Colloquia*. Vol. 15. Liege International Astrophysical Colloquia, p. 59.

Scaramella, R. et al. (Aug. 2021). 'Euclid preparation: I. The Euclid Wide Survey'. In: *arXiv e-prints*, arXiv:2108.01201, arXiv:2108.01201.
arXiv: 2108.01201 (astro-ph.CO).

Schneider, Peter (Dec. 1996). 'Detection of (dark) matter concentrations via weak gravitational lensing'. In: *Monthly Notices of the Royal Astronomical Society* 283.3, pp. 837–853. arXiv: astro-ph/9601039 [astro-ph].

Schneider, Peter (2015). *Extragalactic Astronomy and Cosmology: An Introduction*.

Springel, Volker (Dec. 2005). 'The cosmological simulation code GADGET-2'. In: *Monthly Notices of the Royal Astronomical Society* 364.4, pp. 1105–1134.
arXiv: astro-ph/0505010 (astro-ph).

Springel, Volker, Naoki Yoshida and Simon D. M. White (Apr. 2001a). 'GADGET: a code for collisionless and gasdynamical cosmological simulations'. In: *New Astronomy* 6.2, pp. 79–117. arXiv: astro-ph/0003162 [astro-ph].

Springel, Volker et al. (Dec. 2001b). 'Populating a cluster of galaxies - I. Results at [formmu2]z=0'. In: *Monthly Notices of the Royal Astronomical Society* 328.3, pp. 726–750. arXiv: astro-ph/0012055 [astro-ph].

Taylor, M. B. (Dec. 2005). 'TOPCAT & STIL: Starlink Table/VOTable Processing Software'. In: *Astronomical Data Analysis Software and Systems XIV*. Ed. by P. Shopbell, M. Britton and R. Ebert. Vol. 347. Astronomical Society of the Pacific Conference Series, p. 29.

Taylor, M. B. (July 2006). 'STILTS - A Package for Command-Line Processing of Tabular Data'. In: *Astronomical Data Analysis Software and Systems XV*. Ed. by C. Gabriel et al. Vol. 351. Astronomical Society of the Pacific Conference Series, p. 666.

Umetsu, Keiichi (Dec. 2020). 'Cluster-galaxy weak lensing'. In: *Astronomy and Astrophysics Review* 28.1, 7, p. 7. arXiv: 2007.00506 [astro-ph.CO].

Wright, Candace Oaxaca and Tereasa G. Brainerd (May 2000). 'Gravitational Lensing by NFW Halos'. In: *Astrophysical Journal* 534.1, pp. 34–40.