Methods for Joint Cluster Reconstructions

Korbinian Huber

München 2018

Methods for Joint Cluster Reconstructions

Korbinian Huber

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Korbinian Huber
aus Augsburg

München, den 30. November 2018

Erstgutachter: Prof. Dr. Andreas Burkert
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 13. Februar 2019

"Digressions, incontestably, are the sunshine"

Lawrence Sterne

Zusammenfassung

Galaxienhaufen sind in der modernen Kosmologie von herausragendem Interesse, da sie Schnittpunkte des Netzwerks kosmischer Strukturen sind, und als solche empfindliche Tests von Theorien der Strukturbildung erlauben. Darüberhinaus zeigen sie eine Vielzahl astrophysikalischer Phänomene über weite Bereiche des elektromagnetischen Spektrums, was sie selbst zu interessanten astrophysikalischen Laboren macht. Galaxienhaufen werden dementsprechend in nahezu allen Beobachtungskanälen eingehen untersucht.

In dieser Arbeit präsentiere ich Bemühungen, diese verschiedenen Daten in konsistenter Weise zu kombinieren, um die Massenverteilung innerhalb eines Galaxienhaufens zu rekonstruieren und dabei möglichst wenige Annahmen über selbige zu machen. In Kapitel 1 gebe ich eine kurze Einführung in die kosmologischen Zusammenhänge und die Methoden zur Nutzung von Gravitationslinseneffekten, um dann in Kapitel 2 Galaxienhaufen und ihre Nutzung in Kosmologie und Astrophysik zu diskutieren. In Kapitel 3 behandle ich zunächst Rekonstruktionstechniken im Allgemeinen, bevor ich detailliert unsere gitter- und modellunabhängige Methode erläutere und wie verschiedene Beobachtungsdaten darin eingebunden werden können. Schlussendlich beschreibe ich in Kapitel 4 numerische Testfälle, in denen ich unsere Methode auf simulierte Haufen verschiedenster Komplexität anwende und dabei Daten aus dem schwachen Gravitationslinseneffekt und Röntgenbeobachtungen kombiniere, während sich in Kapitel 5 abschließende Bemerkungen finden.

Die Werkzeuge, die hier entwickelt und vorgestellt werden, erlauben, die Zulässigkeit häufiger Annahmen im Bezug auf das Intraclustermedium, wie polytrope Schichtung und hydrostatischen Gleichgewicht, ebenso zu prüfen wie Annahmen über die Morphologie der zugrundeliegenden Massenverteilung basierend auf numerischen Simulationen. Sie bieten damit die Möglichkeit, systematische Fehlerquellen in einer Vielzahl von Anwendungen, die Galaxienhaufen involvieren, zu untersuchen.

Abstract

Clusters of galaxies are of particular interest in modern Cosmology, as they constitute the signposts of the cosmic web and as such are sensitive probes to test theories of structure formation. They additionally offer a rich astrophysical phenomenology across wavelengths, which makes them interesting astrophysical laboratories in their own right and, consequently, they are studied in great detail in almost all observational channels.

In this work, I present efforts to combine these various data sets to constrain the matter distribution in galaxy clusters in a consistent way, while making as few a priori assumptions on it as possible. In Chapter 1, I set the stage by providing a brief introduction to the cosmological context and the methods of gravitational lensing. In Chapter 2 I discuss galaxy clusters and their uses in Cosmology and Astrophysics. In Chapter 3, I first give an overview of reconstruction methods in general, before detailing our mesh-free and free-form reconstruction framework and how different types of observations can be incorporated in it. Finally, I describe numerical test cases in Chapter 4, where I applied our method to simulated clusters of different complexity, combining weak gravitational lensing and X-ray data, before concluding in Chapter 5.

The tools developed and presented in this work allow us to test the validity of common assumptions on the physics of the ICM like hydrostatic equilibrium and polytropic stratification, as well as assumptions on the morphology of the underlying total mass distribution based on numerical simulations. They thus offer means to investigate systematic uncertainties in a variety of studies that involve clusters of galaxies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This work treads some not very sharp lines. It overlaps with cosmology as well as with astrophysics. Cosmology attempts to understand the laws that govern the evolution of the observable Universe on the largest scales and throughout its entire existence. Astrophysics describes the somewhat more localised, yet still often tremendously large constituents of this Universe, like clusters of galaxies which span several megaparsec, and how they come about. Cosmology forms the context of astrophysics and astrophysics is employed to inform cosmology.
The other ill-defined boundary that is touched upon here is the one between theory and observation. Dealing mostly with the development of numerical methods, neither are any dedicated observations done here, nor is any specific theoretical model developed or tested. The goal is rather to create a set of tools that can be applied to observational data to inform tests of theoretical predictions. Explicitly and in short, the goal of this project was to develop a framework to combine all available data on an individual cluster of galaxies to reconstruct its matter distribution, while making as little assumptions as possible on the latter. Why? Because galaxy clusters are among the main objects of interest in modern cosmology, as they are not only a prominent culmination of structure formation, tracing the evolution of dark matter through cosmic history. They also act as strong gravitational lenses facilitating the study of extremely distant galaxies and as astrophysical laboratories displaying the interplay of radiation and matter in various states and spanning vast scales. All this depends on the knowledge of the properties of a cluster, so a method of cross-checking assumptions on them is of use.

As is customary, I will set the stage by briefly introducing the main concepts of standard cosmology. In doing so, I will mainly keep to the cosmological standard model ΛCDM, which assumes general relativity as the theory of gravity, a non-zero cosmological constant Λ driving the accelerated expansion of the Universe at late times and the existence of a cold dark matter component. I will however highlight possible opportunities for deviations from this concordance model. I will first describe the evolution of the homogeneous background before moving on to the formation of structures embedded in it. Finally I will briefly introduce the basics of gravitational lensing, as it is fundamental to the reconstruction method presented here.

1.1 Background cosmology

The base assumption underlying almost all of modern cosmology is the so-called cosmological principle, which states that on sufficiently large scales, the Universe is homogeneous and isotropic. That means that, if one were to ignore or smooth out everything smaller than roughly 100 Mpc[1], the Universe would look the same, regardless of where you are (*homogeneity*) or in which direction you look (*isotropy*).[2] Applied to a metric theory of gravity, the cosmological principle leads to the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric (Friedmann, 1922; Lemaître, 1927; Robertson, 1935; Walker, 1937) described by the line element[3]

$$\mathrm{d}s^2 = -\mathrm{c}^2\mathrm{d}t^2 + a^2(t)\left[\mathrm{d}\chi^2 + f_\mathrm{K}^2(\chi)\mathrm{d}\Omega^2\right], \tag{1.1}$$

where χ is the comoving radial distance and Ω is the solid angle on the celestial sphere. The isotropy in the cosmological principle enforces spherical symmetry on the spatial part of the metric, which motivates the use of polar coordinates centred on an observer, and the only remaining degrees of freedom are a global scaling of the spatial hypersurface with the scale factor $a(t)$ and the relation between the radial coordinate χ and the area of spheres of constant χ, which is set by the function $f_\mathrm{K}(\chi)$. It is given by

$$f_\mathrm{K}(\chi) = \begin{cases} |K|^{-1/2}\sinh\left(\sqrt{K}\chi\right) & \text{if } K < 0 \\ \chi & \text{if } K = 0\,, \\ K^{-1/2}\sin\left(\sqrt{K}\chi\right) & \text{if } K > 0 \end{cases} \tag{1.2}$$

with K parametrising the spatial curvature.

In general relativity (GR), the evolution of the metric $g_{\mu\nu}$ is given by the Einstein equation (Einstein, 1915; Einstein, 1916)

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi\mathrm{G}}{\mathrm{c}^4} T_{\mu\nu}, \tag{1.3}$$

where the Einstein tensor $G_{\mu\nu}$ encodes the geometric properties of space-time, Λ is the cosmological constant and the stress-energy tensor $T_{\mu\nu}$ describes the matter and energy content of the Universe. Applied to the FLRW metric, and noting that by the cosmological principle, the stress-energy tensor has to be

[1] This is an empirical value, or rather an empirical order of magnitude.

[2] Fortunately this postulate is reasonably well supported by observations as things get very tricky if it gets relaxed. Nonetheless these more general cases are studied, if only to test the cosmological principle (Meyer, Redlich, and Bartelmann, 2015)

[3] The metric signature is chosen to be $(-,+,+,+)$

that of one or more homogeneous perfect fluids with densities ρ_i and pressures p_i, this yields the Friedmann equations

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\sum_i \rho_i - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3}, \tag{1.4}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\sum_i \left(\rho_i + \frac{3p_i}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{1.5}$$

governing the time evolution of the scale factor a. Here we defined the Hubble function H, describing the rate of expansion (or contraction) of the Universe[4]. It is customary to define the scale factor to be unity $a(t_0) = a_0 = 1$ today, such that the Hubble constant $H_0 = \dot{a}_0$ quantifies the expansion rate of the Universe as observed today. Another convention ist to express the Hubble constant as $H_0 = h \cdot 100\,\frac{\mathrm{km/s}}{\mathrm{Mpc}}$ such that the observed value gets absorbed in the dimensionless parameter h.

For any given species, as well as for their sum, the Friedmann equations (1.4) and (1.5) can be combined to obtain the adiabatic equation

$$3H\left(\rho c^2 + p\right) + \dot{\rho}c^2 = 0, \tag{1.6}$$

essentially giving the cosmological version of energy conservation. Integrating Eq. (1.6) and introducing the equation of state parameter

$$w = \frac{p}{\rho c^2}, \tag{1.7}$$

we arrive at an expression for the evolution of the density of a perfect fluid depending on said parameter, assuming it's constant,

$$\rho(a) = \rho_0 a^{-3(1+w)}, \tag{1.8}$$

where ρ_0 again is taken at present day. This allows to distinguish the evolution of different fluid species in an expanding Universe, based on their equation of state. In this context it is convenient to treat spatial curvature and the cosmological constant as fluid species as well and assign them density and pressure in accordance to the Friedmann equations. The most relevant fluids are collected in Table 1.1.

With this and Eq. (1.5) in mind it becomes obvious that a Universe dominated by a non-zero cosmological constant undergoes accelerated expansion ($\ddot{a} > 0$). In fact, every fluid component with $w < -1/3$ would be sufficient to drive this accelerated expansion, which is where cosmological models that attempt to replace the cosmological constant with a dynamical dark energy component set in (see for example Amendola and Tsujikawa, 2010 for a whole book on the efforts along this line). It is worth to briefly note a few things

[4] or equivalently the contraction (or expansion) of our measure of distance

Fluid	EOS parameter	dilution
radiation	$w_\mathrm{r} = 1/3$	$\rho_\mathrm{r} \propto a^{-4}$
(collisionless) matter	$w_\mathrm{m} = 0$	$\rho_\mathrm{m} \propto a^{-3}$
spatial curvature	$w_\mathrm{k} = -1/3$	$\rho_\mathrm{k} = -\frac{3Kc^2}{8\pi G a^2} \propto a^{-2}$
cosmological constant	$w_\Lambda = -1$	$\rho_\Lambda = \frac{\Lambda c^2}{8\pi G} \propto$ const.

TABLE 1.1: Collection of relevant fluid species, their equation of state (EOS) parameters and their evolution with the scale factor

here. From the view point of GR, there is absolutely no problem with a cosmological constant $\Lambda > \frac{4\pi G}{c^2}\left(\rho + \frac{3p}{c^2}\right)$, driving the accelerated expansion. In fact, GR with a cosmological constant is the most general metric theory of gravity in four dimensions with equations of motion that are linear in the second order derivatives of the metric (Lovelock, 1972). Hence the presence of Λ is expected and its value, as any constant of nature, is up to measurements. The ever-growing field of dark energy and modified gravity theories owes its existence to problems that arise when attempting to reconcile GR and Quantum Field Theory (QFT), as the measured value of Λ is vastly different from QFT estimates of vacuum energy and its gravitational effect. Attempts to overcome this discrepancy from the side of cosmology are usually categorized into modified gravity or dark energy theories, depending on wether they alter the underlying assumptions of gravity or introduce new fluids to the stress energy tensor.

The first Friedmann equation (1.4) hints at a characteristic density scale

$$\rho_\mathrm{crit}(t) = \frac{3H^2(t)}{8\pi G} \tag{1.9}$$

called the critical density, as it sets the threshold for the geometry of the Universe. If the densities of matter, radiation and the cosmological constant add up to a value above ρ_crit, the spatial curvature becomes positive, leading to spherical geometry, whereas a value below indicates negative curvature and hyperbolic geometry. Only if all fluid densities add up to the critical density exactly, the spatial curvature vanishes and we are left with flat Euclidean geometry. With the critical density, or rather its present day value $\rho_\mathrm{crit,0} \equiv \rho_\mathrm{crit}(t = t_0)$, we can rewrite Eq. (1.4) in terms of dimensionless density parameters

$$\begin{aligned} H^2(a) &= H_0^2 \left(\frac{\Omega_\mathrm{r,0}}{a^4} + \frac{\Omega_\mathrm{m,0}}{a^3} + \frac{\Omega_\mathrm{k,0}}{a^2} + \Omega_\mathrm{\Lambda,0} \right) \\ &= H_0^2 \left(\Omega_\mathrm{r}(a) + \Omega_\mathrm{m}(a) + \Omega_\mathrm{k}(a) + \Omega_\Lambda \right) \qquad (1.10) \\ &\equiv H_0^2 E^2(a), \qquad (1.11) \end{aligned}$$

where

$$\Omega_{i,0} = \frac{\rho_{i,0}}{\rho_{\mathrm{crit},0}}. \tag{1.12}$$

Thus the expansion history of the Universe can be recovered from the values of these Ω-parameters today and vice versa. The density parameters (including $\Omega_{\mathrm{k},0}$) add up to 1 at any time by construction. As the data indicates a flat Universe now (Planck Collaboration et al., 2016) and since the importance of spatial curvature relative to matter and radiation increases in a decelerating Universe[5], $|\Omega_\mathrm{k}|$ either was vanishingly small initially or was reduced by a previous phase of accelerated expansion, which is one motivation for cosmic inflation (Baumann, 2007). After this follow successively and according to their dilution behaviour, phases of radiation domination, matter domination and recently[6] a phase of accelerated expansion dominated by the cosmological constant.

As the metric gets scaled during the expansion of the Universe, the wavelength of photons gets shifted towards higher values. This cosmological redshift is given by

$$z \equiv \frac{\lambda_\mathrm{obs} - \lambda_\mathrm{em}}{\lambda_\mathrm{em}}, \tag{1.13}$$

where λ_obs is the observed wavelength of a photon and λ_em is its wavelength when emitted. The cosmological redshift relates to the scale factor via

$$z = \frac{1-a}{a} \tag{1.14}$$

and provides a convenient way to determine the latter at the time a given light source emitted the photons observed today (assuming knowledge on the original wavelength). In order to reconstruct for example the expansion history from these measurements, we have to assess the distance to these sources as well. Thus, measuring distances is one of the key tasks in cosmology and several distance measures have been introduced, as there is no unique notion of distance in a dynamical and in general curved spacetime.

The *comoving distance* is the distance on a spatial hypersurface of constant time (see Eq. 1.1). It remains unaffected by the dynamics of the background, hence the name. As a function of redshift, the comoving distance between an observer and a light emitting source is given by

$$\chi(z_\mathrm{obs}, z_\mathrm{em}) = \int_{a(z_\mathrm{obs})}^{a(z_\mathrm{em})} \frac{\mathrm{d}a}{a\dot{a}} = \frac{c}{H_0} \int_{a(z_\mathrm{obs})}^{a(z_\mathrm{em})} \frac{\mathrm{d}a}{a^2 E(a)}. \tag{1.15}$$

[5] the deviation from flatness is proportional to the Hubble radius which only shrinks during accelerated expansion

[6] cosmically speaking; the cosmological constant started dominating about 3.6 billion years ago

Parameter	value
$\Omega_{\mathrm{m},0}$	0.3089 ± 0.0062
$\Omega_{\Lambda,0}$	0.6911 ± 0.0062
$\Omega_{\mathrm{k},0}$	$0.0008^{+0.0040}_{-0.0039}$
$\Omega_{\mathrm{r},0}$	$(5.93 \pm 0.07) \cdot 10^{-5}$
H_0	67.74 ± 0.46

TABLE 1.2: Recent values of cosmological parameters according to (Planck Collaboration et al., 2016); based on measurements of the Cosmic Microwave Background (CMB) in combination with external probes (baryon acoustic oscillation, supernovae type Ia, H_0 based on Cepheids).

The *proper distance* is the distance measured by the travel time of a photon between source and observer and can be expressed as

$$D_{\mathrm{prop}}(z_{\mathrm{obs}}, z_{\mathrm{em}}) = \int_{a(z_{\mathrm{obs}})}^{a(z_{\mathrm{em}})} \frac{\mathrm{d}a}{\dot{a}} = \frac{c}{H_0} \int_{a(z_{\mathrm{obs}})}^{a(z_{\mathrm{em}})} \frac{\mathrm{d}a}{aE(a)}. \tag{1.16}$$

The *angular diameter distance* relates the intrinsic size δA of an object to the solid angle $\delta\Omega$ under which it appears for an observer. It is given by

$$D_{\mathrm{ang}}(z_{\mathrm{obs}}, z_{\mathrm{em}}) = \frac{\delta A}{\delta\Omega} = \frac{1 + z_{\mathrm{obs}}}{1 + z_{\mathrm{em}}} f_{\mathrm{K}} \left[\chi(z_{\mathrm{obs}}, z_{\mathrm{em}})\right]. \tag{1.17}$$

Finally, the *luminosity distance* relates the intrinsic luminosity L of a light source to its observable flux F and is defined as

$$D_{\mathrm{lum}}(z_{\mathrm{obs}}, z_{\mathrm{em}}) = \sqrt{\frac{L}{4\pi F}} = \left(\frac{1 + z_{\mathrm{em}}}{1 + z_{\mathrm{obs}}}\right)^2 D_{\mathrm{ang}}(z_{\mathrm{obs}}, z_{\mathrm{em}}). \tag{1.18}$$

In practice only the latter two are directly observable and which distant measure should be used entirely depends on the situation at hand. So, if we know the intrinsic luminosities or sizes of a set of sources, we can determine their distance and, by an independent measurement of their redshift via spectroscopy, infer the expansion history and consequently constrain the cosmological parameters. Recent values on these and some other parameters are collected in Table 1.2. If not stated otherwise, they constitute the cosmological model adopted in this work.

1.2 Structure

Since we seem to exist, just as stars, galaxies and the cosmic large scale structure do, there clearly are pronounced inhomogeneities in the Universe and

the cosmological principle is broken on scales below roughly 100 Mpc. The following section will deal with evolution of inhomogeneities in the Universe.

1.2.1 Dark matter

On scales on which the cosmological principle holds, gravity is the only relevant force, so the behaviour of the components of the stress-energy tensor with respect to the other fundamental forces is of no interest. Therefore all non-relativistic matter can be treated as what GR literature refers to as dust, that is collisionless matter ($w = 0$). This changes when matter is allowed to stream and cluster and when it is supposed to be observed as observation implies at least some interaction with light[7]. The cosmological terminology distinguishes between *baryonic matter*, essentially comprised of all the massive particles in the standard model of particle physics that interact with the electromagnetic force, and *dark matter*, which only interacts gravitationally (and at most via the weak force). Note that this means that massive neutrinos are in fact an example of dark matter, albeit a subdominant one when it comes to cosmology. Dark matter was initially postulated to account for apparent differences in gravitational and stellar mass of the Coma cluster (Zwicky, 1933) and for the flatness of galactic rotation curves on larger radii than expected (Rubin, Ford, and Thonnard, 1978). Nowadays however, the most powerful evidence for the existence of dark matter comes from the cosmic microwave background (CMB). Current data requires a mass component that is non-baryonic, mostly collisionless, does not interact with light (except gravitationally) and has a very small or negligible velocity dispersion (Planck Collaboration et al., 2016). The latter is what distinguishes *cold dark matter* (CDM), where the intrinsic velocity dispersion is negligible compared to the speed of light, from *hot dark matter* (HDM) that travels at relativistic speeds. Neutrinos seem to fall in the latter category, as their masses are low and they decouple at high enough temperatures to retain relativistic velocities (Lesgourgues and Pastor, 2006). Cold dark matter forms structures bottom-up, where low mass objects form first, accrete more matter and merge with each other to form higher mass objects, whereas hot dark matter leads to a top-down structure formation where high mass objects form first (and might fragment further) and the free streaming of the HDM particles prevents the formation of structures on galactic scales, a scenario that is essentially ruled out by data (White, Frenk, and Davis, 1983). *Warm dark matter* (WDM) lies somewhere in between, so structures form essentially bottom up but those below a certain mass scale are washed out by the free streaming of the WDM particles.

Theoretical particle physics beyond the standard model provides a plethora of potential candidates for dark matter and I will only mention a few. Weakly interacting massive particles (WIMPs) have long been considered to be good candidates, partly since these thermal relics with masses in the range of 100 GeV and small self-interaction cross sections are supported by for example the theory of supersymmetry and would be produced in approximately

[7] with the very recent exceptions of gravitational wave and neutrino astronomy

the correct abundance[8](Bertone, 2010). As significant efforts to directly or indirectly detect WIMPs so far only have yielded negative results (Marrodán Undagoitia and Rauch, 2016), just as searches for signs of supersymmetry in collider experiments (Buchmueller et al., 2014), support for WIMPs as the main contribution to cold dark matter starts to cease. Lighter thermal relics in the keV to GeV range are considered candidates for WDM.
Axions (Weinberg, 1978; Visinelli and Gondolo, 2009) are the predicted particle consequence of a suggested mechanism to solve the strong CP problem of quantum chromodynamics (QCD), which deals with the question why experiments do not show any evidence of a violation of CP symmetry in the strong interaction, even though it would be allowed by QCD. If axions existed, they would display many features of cold dark matter, specifically no electric charge and small self-interaction cross section.

While collisionless particle dark matter is the framework favoured by the literature nowadays, it is not the only one. Dark matter models with non-vanishing self-interaction or couplings to dark energy (or even a whole dark sector) have been proposed (Spergel and Steinhardt, 2000; Farrar and Peebles, 2004). Other models for dark components include massive compact halo objects (MACHOs) like black holes or very faint stellar remnants. Microlensing studies did however not find sufficient amounts of such objects for them to be a likely alternative (Alcock et al., 2000)[9] and they wouldn't meet the criterion of being at most gravitationally and weakly interacting. Outside the dark matter framework, there is a variety of modified gravity theories attempting to explain the observational features ascribed to dark matter, especially in galaxies, by deviations from GR rather than an unknown particle, e.g. Bekenstein (2004). These currently fail to accommodate more than but a few specialised observations, though (Mavromatos, Sakellariadou, and Yusaf, 2009; Reyes et al., 2010).

In the following description of structure formation, a standard collisionless particle cold dark matter framework is assumed. The reconstruction method developed in this work presupposes a sufficiently smoothly distributed dark matter component dominating the gravitational potential of galaxy clusters, but no assumptions on a specific type or particle are made.

1.2.2 Linear inhomogeneities

The central quantity for the linear evolution of structure is the density contrast

$$\delta(\boldsymbol{x},t) = \frac{\rho(\boldsymbol{x},t) - \langle\rho\rangle(t)}{\langle\rho\rangle(t)}, \tag{1.19}$$

where $\langle\rho\rangle(t)$ is the mean matter density in the Universe. We continue to treat the density contrast as that of a fluid and apply the hydrodynamical equations

[8] this is referred to as the WIMP miracle

[9] although the degree to which candidates like primordial black holes are ruled out is up to debate

to account for its evolution, even though dark matter is assumed to be collisionless and therefore has a non-negligible mean free path length, making hydrodynamics in principle ill-equipped to describe it. Since we are dealing with the linearised evolution in this section however, the hydrodynamical formalism remains a good approximation. In particular, we will employ the Euler equation

$$\dot{\boldsymbol{v}} + (\boldsymbol{v}\nabla)\boldsymbol{v} = -\nabla p - \nabla\Phi, \tag{1.20}$$

describing the evolution of the peculiar velocity field $\boldsymbol{v}$ in presence of a pressure gradient and gravitational forces, the continuity equation

$$\dot{\rho} + \nabla(\rho\boldsymbol{v}) = 0 \tag{1.21}$$

and the Poisson equation

$$\nabla^2\Phi = 4\pi G\rho. \tag{1.22}$$

Restricting ourselves to small perturbations δ, $\boldsymbol{v} \ll 1$, δp and $\delta\Phi$ and transforming to comoving coordinates

$$\boldsymbol{x} \to \boldsymbol{x}/a, \quad \boldsymbol{v} \to \boldsymbol{v} + H\boldsymbol{x}, \tag{1.23}$$

the above equations take the form

$$\dot{\boldsymbol{v}} + H\boldsymbol{v} = -\frac{1}{a\langle\rho\rangle}\nabla\delta p - \frac{1}{a}\nabla\delta\Phi, \tag{1.24}$$

for the Euler equation,

$$\dot{\delta} = -\frac{1}{a}\nabla\boldsymbol{v} \tag{1.25}$$

for the continuity equation and

$$\nabla^2\delta\Phi = 4\pi G a^2\langle\rho\rangle\delta \tag{1.26}$$

for the Poisson equation. Combining these linearised versions of the fluid equations finally yields the *linear growth equation*

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\langle\rho\rangle\delta + \frac{w}{a^2}\nabla^2\delta, \tag{1.27}$$

where $w = \delta p/\delta\rho$ is the equation of state parameter. Since for linearly evolving pressureless matter on cosmic scales $w = 0$, the last term can be dropped. Equation (1.27) has a growing solution $D_+(a)$ such that

$$\delta(a) = D_+(a)\delta_0. \tag{1.28}$$

The exact form of this growth function D_+ depends on the background cosmology via the Hubble function in the friction term in the growth equation and the mean matter density $\langle\rho\rangle$. While linear growth of (dark) matter perturbations is suppressed in the epoch of radiation domination ($D_+ \propto \ln a$) and does not happen at all in the de-Sitter case ($\Omega_\Lambda = 1$), perturbations grow with $D_+ \propto a$

during matter domination, so most of the linear growth enabling the formation of structures we see today has to have happened in that epoch.

Note that the approach above is entirely Newtonian, an approximation that is justifiable only because we assume linearity, sub-horizon scales and small perturbations $\delta \ll 1$. These assumptions are also responsible for the fact that the growth function is not a function of position or perturbation scale but only of the background scale factor. As soon as growth becomes non-linear, this changes.

1.2.3 Statistics of inhomogeneities

In a standard inflationary model, the primordial inhomogeneities (or equivalently anisotropies, if we restrict ourselves to one observer's position) in the Universe are sourced by quantum fluctuations in the density field that get expanded to macroscopic scales during inflation and subsequently evolve under their own gravity and influenced by the cosmological background evolution. This quantum origin leads to the prediction[10] that the primordial anisotropies are a Gaussian random field (Baumann, 2007), a view that is fully supported by current observations (Planck Collaboration et al., 2016). As it is the Gaussian way (due to Wick's theorem), the statistics of this random field are completely described by the mean an the (co-) variance. The former is zero by construction as we are dealing with the density contrast, while the latter is given by the correlation function

$$\xi(\boldsymbol{x}_1, \boldsymbol{x}_2) = \langle \delta(\boldsymbol{x}_1)\delta(\boldsymbol{x}_2)\rangle \,, \tag{1.29}$$

where the averages are taken over a virtual ensemble of realisations[11]. The density field is further assumed to be statistically homogeneous and isotropic, meaning that the correlation function does not actually depend on the position or relative orientation of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, but only on their relative distance $r = |\boldsymbol{r}| = |\boldsymbol{x}_2 - \boldsymbol{x}_1|$,

$$\xi(r) = \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle \,. \tag{1.30}$$

It is convenient to transform the density contrast to Fourier space and calculate the covariance there

$$\langle \delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle = (2\pi)^3 \delta_{\mathrm{D}}(\boldsymbol{k}_1 - \boldsymbol{k}_2) P(k), \tag{1.31}$$

where we defined the power spectrum $P(k)$ and the Dirac delta function expresses statistically homogeneous Gaussianity by ensuring that different modes do not couple. The power spectrum relates to the real space 2-point correlation function via the Fourier transformation

$$\xi(r) = 4\pi \int \frac{k^2 \mathrm{d}k}{(2\pi)^3} P(k) \frac{\sin(kr)}{kr} = 4\pi \int \frac{k^2 \mathrm{d}k}{(2\pi)^3} P(k) j_0(kr), \tag{1.32}$$

[10] shared by many inflationary models, but not necessarily all

[11] As this is not possible in practice, one usually takes averages over causally disconnected portions of the sky, introducing an irreducible uncertainty called cosmic variance.

with the spherical Bessel function $j_0(x)$. Again, inflationary models make predictions for the initial power spectrum to be a nearly scale invariant power law

$$P(k) \propto k^{n_\mathrm{s}} \tag{1.33}$$

with spectral index $n_\mathrm{s} \lesssim 1$ (Mukhanov and Chibisov, 1981). Again, measurements so far confirmed this. The power spectrum is customarily normalized such that the variance of the density contrast averaged over a certain radius R

$$\sigma_R^2 = 4\pi \int \frac{k^2 \mathrm{d}k}{(2\pi)^3} P(k) W_R^2(k) \tag{1.34}$$

is fixed to an observed value, by convention for $R = 8\,\mathrm{Mpc}/h$. Here

$$W_R(k) = \frac{3 j_1(kR)}{kR} \tag{1.35}$$

is the Fourier transform of a spherical top-hat filter of radius R in real space, expressed using the spherical Bessel function $j_1(x)$.

1.2.4 The Cosmic Microwave Background

The CMB has been mentioned before in this chapter, but given its paramount importance for modern cosmology and its peripheral relevance for this work, it is worthwhile to spend a few more words on it. For roughly the first 300 000 years after the Big Bang[12] the baryonic matter content of the Universe remained in the state of an ionized hydrogen plasma (with contaminations of heavier elements as they were produced by BBN[13] mechanisms). In this state the matter was tightly coupled to the radiation component and the mean free path of photons remained small. As the Universe expanded, the plasma cooled adiabatically and remained in thermal equilibrium with photons until the temperature was low enough for electrons and protons to recombine and form neutral hydrogen. As this process of recombination happened relatively sudden[14], the photons quickly decoupled from the baryonic matter, retaining a black body spectrum. During their relatively unhindered travel since recombination at $z \approx 1100$, they get redshifted accordingly, such that nowadays we observe this radiation as a very isotropic microwave background showing an almost perfect Planck spectrum with $T_\mathrm{CMB} \approx 2.725\,\mathrm{K}$ and only tiny anisotropies of order $\delta T \sim 10^{-5} T_\mathrm{CMB}$. These temperature fluctuations, depicted in fig. (1.1), relate to the density fluctuations described above via various processes. On large angular scales the Sachs-Wolfe effect (Sachs and Wolfe, 1967) dominates, roughly stating that gravitational redshift causes photons released from overdense, intrinsically hotter regions to appear cooler than those

[12] or the end of inflation

[13] Big Bang nucleosynthesis, i.e. the formation of elements heavier than hydrogen and lighter than beryllium during the radiation dominated era

[14] again, cosmically speaking; the width of the recombination shell in redshift is well approximated by a Gaussian with mean $z \approx 1100$ and standard deviation $\sigma_z \approx 80$.

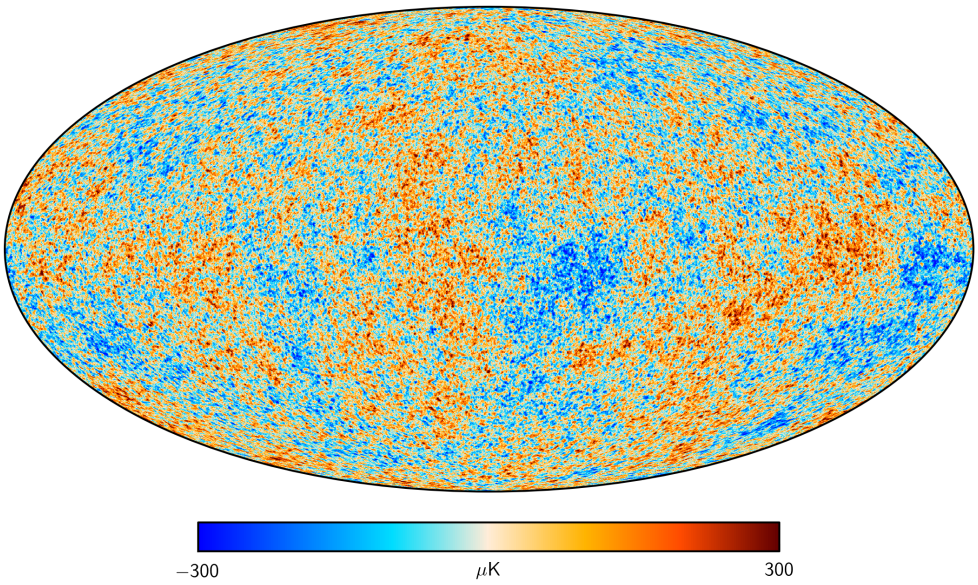

FIGURE 1.1: Foreground corrected and interpolated map of the CMB temperature fluctuations based on (Planck Collaboration et al., 2016)

from underdense regions. On scales smaller than the sound horizion at recombination, baryonic accoustic oscillations, i.e. sound resonances in the density field are the predominant source of temperature fluctuations. The angular power spectrum of these oscillations in turn is sensitive to the relative prevalence of the cosmic fluid species. Among others, temperature measurements of the CMB can be used to (Hu and Dodelson, 2002)

- constrain the spatial curvature via the sound horizon as a standard ruler, provided an independent measurement of H_0
- infer the existence of an electromagnetically non-interacting matter species from the amplitude of baryonic density fluctuations that would be insufficient to account for the large density contrast observed today
- infer the relative prevalences, i.e. the Ω-parameters, of radiation, baryonic and dark matter at time of recombination

1.2.5 Non-linear evolution

The description of growth as given in section 1.2.2 is an example of Eulerian perturbation theory, tracing the evolution of small perturbations to the density and velocity fields. Another approach would be Lagrangian perturbation theory, which follows the trajectories of fluid elements by evolving the field of their displacements. Either approach breaks down, though, once the density contrast reaches values of order unity. As an overdensity grows, the local gravitational potential well steepens, therefore altering the local growth equation. This means that the growth function becomes position dependent, leading to

mode coupling in a Fourier space formalism. As the density contrast deviates further and further from the mean, its statistics become increasingly skewed and non-Gaussian, signifying that simple 2-point correlation functions are no longer sufficient to fully describe the density field.

There is significant and ongoing effort to extent the range of Eulerian and Lagrangian perturbation theories to the non-linear regime (Bernardeau et al., 2002) but so far with limited success. Alternative approaches, for example based on a kinetic field theory of correlated many-body systems are being developed (Bartelmann et al., 2016) but so far no analytic method can fully describe the phenomenology of non-linear structure formation and one has to resort to numerical simulations (Bertschinger, 1998; Springel, 2010).

Analytical approximations can however still give useful insights as we shall see on the example of *spherical collapse* (Weinberg, 2008). This approximation starts out from the assumption of a uniform overdensity confined to a finite sphere. The evolution of such a perturbation by Birkhoff's theorem decouples from the background and can effectively be treated as a separate, closed ($K > 0$) Universe. Its radial extent is governed by

$$\ddot{R} = -\frac{GM}{R^2}, \tag{1.36}$$

where M is the total mass enclosed in R. Integration of this equation just recovers the energy of the system

$$\frac{1}{2}\dot{R}^2 - \frac{GM}{R} = E. \tag{1.37}$$

The collapsing case ($E < 0$) is solved in parametric form by

$$R = A\,(1 - \cos\theta) \tag{1.38}$$
$$t = B\,(\theta - \sin\theta) \tag{1.39}$$
$$A^3 = GMB^2, \tag{1.40}$$

where the state of the system is parametrised by the phase θ. With this we immediately see that the radial extent grows initially before turning around at $\theta = \pi$ and finally re-collapsing to, formally, infinite density at $\theta = 2\pi$. Of course in reality the system does not collapse to a singularity and will rather virialise to a meta-stable equilibrium state of finite radius $R_{\rm vir}$. Nevertheless, the description above offers an estimate on the time scales of collapse. Specifically we can express the density contrast as a function of time, restricting ourselves to linear approximations ($\delta \ll 1$)

$$\delta_{\rm lin.} \approx \frac{6}{20}\left(\frac{6t}{B}\right)^{2/3}. \tag{1.41}$$

Inserting the formal time of collapse $t_{\rm c} = 2\pi B$ yields

$$\delta_{\rm c} \equiv \delta_{\rm lin.}(t_{\rm c}) \approx 1.686, \tag{1.42}$$

where we have defined δ_c as the linearly extrapolated density contrast at time of collapse. Once linear growth equations lead to a contrast of δ_c, we can expect that the system has already collapsed non-linearly and formed a meta-stable halo, with a non-linear virial overdensity of $\Delta \approx 178$[15].

In order to get the abundance of these halos, we consider the smoothed density contrast δ_R, where R is the smoothing scale of, say, a spherical top-hat window function in Fourier space (see section 1.2.3). We can thus compute the cumulative probability for the linear smoothed density to lie above δ_c

$$F(R;\delta_c) = \int_{\delta_c}^{\infty} \mathrm{d}\delta_R p(\delta_R), \tag{1.43}$$

where the probability distribution for the the linear smoothed density contrast is, as stated in section 1.2.3, given by a Gaussian

$$p(\delta;R) = \frac{1}{\sqrt{2\pi}\sigma_R} \exp\left(-\frac{\delta_R^2}{2\sigma_R^2}\right). \tag{1.44}$$

As the mass inside a spherical top-hat overdensity of radius R in real space is given by $M = 4\pi\Omega_m\rho_{crit}R^3/3$, we can also express the cumulative probability as a function of mass[16]

$$F(M) = \int_{\delta_c}^{\infty} \mathrm{d}\delta_m p(\delta_M) = \frac{1}{2}\mathrm{erfc}\left(\frac{\delta_c}{\sqrt{2\sigma_R^2}}\right). \tag{1.45}$$

The abundance of collapsed halos as a function of their mass is then

$$\frac{\partial F(M)}{\partial M} = \frac{1}{\sqrt{2\pi}}\frac{\delta_c}{\sigma_R}\exp\left(-\frac{\delta_c^2}{2\sigma_R^2}\right)\frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M}. \tag{1.46}$$

This however is not normalised correctly as can be seen by integrating over all masses

$$\int_0^{\infty} \mathrm{d}M \frac{\partial F(M)}{\partial M} = \frac{1}{2}. \tag{1.47}$$

The reasons for this apparent miscounting of objects are a bit intricate and are presented in detail for example in Bond et al. (1991). Essentially, if we step by step reduce the smoothing length R in Eq. (1.44) starting from some large value, the density field at a given point describes a random walk, approaching the threshold δ_c. As the Gaussian is symmetric, it is at every step equally likely to go up or down in density. A trajectory that has once crossed the

[15] This is based on estimating the non-linear density at turnaround and taking into account that due to the virial theorem the system has to be eight times denser than that at virialisation whereas the background density got diluted approximately by a factor of 4 in the meantime, assuming a matter dominated universe (results differ for different cosmologies).

[16] Note that, as the window function used for the smoothing is a top-hat in Fourier space and not in real space, this is not entirely rigorous.

threshold has thus even chances to further increase in density or to decrease again below δ_c. As a once collapsed structure in reality cannot un-collapse again, this leads to miscounting and needs to be compensated by a factor of 2. We arrive therefore at the Press-Schechter mass function (Press and Schechter, 1974)

$$n(M) \equiv 2\frac{\partial F(M)}{\partial M} = \sqrt{\frac{2}{\pi}}\frac{\delta_c}{\sigma_R}\exp\left(-\frac{\delta_c^2}{2\sigma_R^2}\right)\frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M}. \tag{1.48}$$

The mass function is characterised by an exponential sensitivity to σ_R^2. In the above derivation we have suppressed the redshift dependence of the variance $\sigma_R^2 \propto D_+(z)$. Together with the cosmological dependence via the power spectrum, this turns the halo mass function into a viable probe of structure formation and cosmology. The Press-Schechter mass function in this form is not a very good description of the real or simulated Universe, though, as the assumptions of spherical collapse are not really met. The real space configuration of dark matter halos is rarely spherical and does in general not correspond to a Fourier or real space top-hat. There have however been several adjustments and corrections accounting for these effects (Sheth, Mo, and Tormen, 2001) and the main concepts surrounding the mass function remain, i.e. the exponential cut-off for high masses and the universal shape, depending more on the relative peak height δ_c/σ_R than directly on cosmology. This is also true for alternative definitions of the mass function that are directly based on numerical simulations, rather than analytical estimations (Tinker et al., 2008; Jenkins et al., 2001).

1.3 Gravitational lensing

The previous sections introduced inhomogeneities in the cosmic density field and described their evolution. These perturbations lead to numerous observational features, among which gravitational lensing might be the most important for this work, as it will define the target and general structure of the reconstruction framework in Chapter 3.

1.3.1 General derivations and terminology

Gravitational lensing describes the deflection of light rays due to metric perturbations along their path. In the following, I will briefly summarize the derivations of the equations and quantities relevant for this work. For more thorough treatments of the subject, see for example Bartelmann and Schneider (2001), Schneider, Kochanek, and Wambsganss (2006), and Bartelmann (2010) on which this section is largely based. For a slightly different approach employing Fermat's principle, see Blandford and Narayan (1986).
Throughout this section it's assumed that the sources of gravitational lensing effects are slow moving (much slower than the speed of light) and have small Newtonian potentials ($\Phi \ll c^2$), i.e. only terms up to first order in the

Newtonian potential will be kept. Starting from a metric theory of gravity, the perturbed FLRW metric in Newtonian gauge reads

$$\mathrm{d}s^2 = -\left(1+\frac{2}{c^2}\Phi\right)c^2\mathrm{d}t^2 + \left(1+\frac{2}{c^2}\Psi\right)\left(\mathrm{d}\chi^2 + f_\mathrm{K}^2(\chi)\mathrm{d}\Omega^2\right), \tag{1.49}$$

where Φ and Ψ are the Bardeen potentials (Bardeen, 1980) and χ and f_K are already familiar from Eq. (1.1). In this gauge, the Bardeen potential Φ in the time-time part of the line element is identical to the Newtonian potential. Photons follow null geodesics ($\mathrm{d}s = 0$), so their propagation is described by the geodesic equation

$$\frac{\mathrm{d}^2x^\mu}{\mathrm{d}\lambda^2} = -\Gamma^\mu_{\alpha\beta}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda} \tag{1.50}$$

with the affine parameter λ parametrising the progression along the trajectory and the Christoffel symbols $\Gamma^\mu_{\alpha\beta}$ giving the metric connection. Identifying the affine parameter with the comoving distance and evaluating the spatial part of the geodesic equation for the given perturbed metric yields

$$\frac{\mathrm{d}^2x^i}{\mathrm{d}\lambda^2} = \frac{\mathrm{d}^2x^i}{\mathrm{d}\chi^2} = \frac{1}{c^2}\partial^i(\Phi-\Psi) \overset{\mathrm{GR}}{=} -\frac{2}{c^2}\partial^i\Phi. \tag{1.51}$$

Only in the last equality general relativity as a specific metric theory of gravity and the absence of anisotropic stress is assumed, such that the Einstein equations imply the identity $\Psi = -\Phi$. This differential equation can now be integrated under the appropriate boundary conditions

$$x^i\Big|_{\chi=0} = 0, \qquad \frac{\mathrm{d}x^i}{\mathrm{d}\chi}\Big|_{\chi=0} = \theta^i \tag{1.52}$$

to get

$$x^i(\chi) = f_\mathrm{K}(\chi)\theta^i - 2\int_0^\chi \mathrm{d}\chi' f_\mathrm{K}(\chi-\chi')\partial^i\Phi\left(x^j(\chi'),\chi'\right). \tag{1.53}$$

Note that the potential gradient in the integrand depends on the perturbed path $x^i(\chi)$ itself. This implicit structure is often circumvented by applying the Born approximation $x^i(\chi) \approx f_\mathrm{K}(\chi)\theta^i$, i.e. assuming small deflection angles and approximating the curved path with straight lines. With this, and after defining angular coordinates at the distance of a light emitting source $x^i(\chi_s) = f_\mathrm{K}(\chi_s)\beta^i$ one finally arrives at the *lens equation*

$$\beta^i = \theta^i - \underbrace{2\int_0^{\chi_\mathrm{s}}\mathrm{d}\chi'\frac{f_\mathrm{K}(\chi_\mathrm{s}-\chi')}{f_\mathrm{K}(\chi_\mathrm{s})}\partial^i\Phi\left(f_\mathrm{K}(\chi')\boldsymbol{\theta},\chi'\right)}_{\alpha^i(\boldsymbol{\theta})} \tag{1.54}$$

$$\Rightarrow \boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{1.55}$$

where we have defined the components of the *deflection angle*[17] $\boldsymbol{\alpha}$.

Introducing derivatives with respect to angular coordinates $\partial^i_\theta = f_\mathrm{K}(\chi)\partial^i$, the deflection angle can be rephrased as the angular gradient of an effective *lensing potential* ψ, which is given by the weighted projection of the Newtonian gravitational potential

$$\psi(\boldsymbol{\theta}) = \frac{2}{c^2}\int_0^{\chi_\mathrm{s}} \mathrm{d}\chi' \frac{f_\mathrm{K}(\chi_\mathrm{s}-\chi')}{f_\mathrm{K}(\chi_\mathrm{s})f_\mathrm{K}(\chi')}\Phi\left(f_\mathrm{K}(\chi')\boldsymbol{\theta},\chi'\right) \tag{1.56}$$

$$\alpha^i(\boldsymbol{\theta}) = \partial^i_\theta\psi(\boldsymbol{\theta}) \equiv \psi_{,i}. \tag{1.57}$$

In the last equation we also defined another commonly used short-hand notation to denote partial derivatives with respect to angular coordinates.
Equations (1.55) and (1.56) completely define a mapping between the celestial sphere of an observer and the sky position of the light emitting source. Photons which are observed at sky coordinates $\boldsymbol{\theta}$ were emitted at $\boldsymbol{\beta}$ and the relation between those is set by the lensing potential.

The Jacobian of this mapping is

$$\mathcal{A}(\boldsymbol{\theta}) \equiv \frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\theta}} = \left(\delta_{ij} - \frac{\partial^2\psi(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right) = \begin{pmatrix}1-\kappa & 0\\ 0 & 1-\kappa\end{pmatrix} - \begin{pmatrix}\gamma_1 & \gamma_2\\ \gamma_2 & -\gamma_1\end{pmatrix}, \tag{1.58}$$

where in the last equality we separated the trace from the trace-free part. The former with the *convergence*

$$\kappa = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial\theta_1^2} + \frac{\partial^2\psi}{\partial\theta_2^2}\right) = \frac{1}{2}\left(\psi_{,11}+\psi_{,22}\right) = \frac{1}{2}\Delta_\theta\psi \tag{1.59}$$

describes isotropic changes in apparent size of lensed images, whereas the trace-free part with the two components of the *complex shear*

$$\gamma = \gamma_1 + \mathrm{i}\gamma_2 \tag{1.60}$$

$$\gamma_1 = \frac{1}{2}\left(\psi_{,11}-\psi_{,22}\right) \tag{1.61}$$

$$\gamma_2 = \psi_{,12} \tag{1.62}$$

describes anisotropic distortions. Phrasing the shear as a complex number further emphasizes that the trace-free part of the Jacobian (the shear matrix) transforms like a field of spin 2 under rotations whereas the convergence (a scalar) remains rotationally invariant (spin 0). Written in polar form $\gamma = |\gamma|\,\mathrm{e}^{2i\phi}$, the amplitude of the shear is indicative of the strength of the distortion and the phase gives its direction.

[17] This quantity is sometimes also referred to as the reduced deflection angle, due to a different derivation of the lens equation

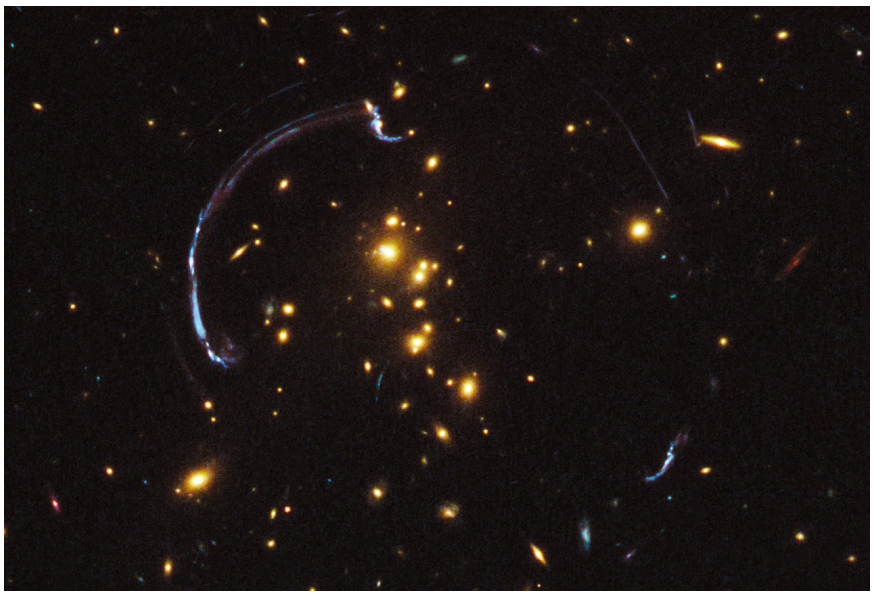

FIGURE 1.2: The galaxy cluster RCS2 032727-132623 prominently features a giant arc caused by strong gravitational lensing of a background galaxy. Picture taken with the Hubble Space Telescope's Wide Field Camera 3[18]

For a circular source of unit radius, the lens mapping leads to an elliptical lensed image with semi-major and semi-minor axes

$$A = (1 - \kappa - |\gamma|)^{-1}, \qquad B = (1 - \kappa + |\gamma|)^{-1}, \tag{1.63}$$

such that the resulting images' ellipticity can be given as

$$\epsilon = \frac{A - B}{A + B} = \frac{|\gamma|}{1 - \kappa} \equiv g, \tag{1.64}$$

which defines the so called *reduced shear* g.

The mapping described by Eq. (1.55) conserves spectral properties and specific intensity (in absence of scattering and absorption) but changes the solid angle under which extended sources appear by the *magnification*

$$\mu = |\det(\mathcal{A})|^{-1} = |(1 - \kappa)^2 - \gamma_1^2 - \gamma_2^2|^{-1}. \tag{1.65}$$

The mapping can become singular ($\det(\mathcal{A}) = 0$) on smooth and closed, so-called *critical curves* on the observer's sky. Their corresponding closed curves at the source distance are called *caustics*. For points within a caustic, the mapping is non-linear, and sources at these positions end up with multiple images on the observers sky. Sources on caustics are extremely magnified (though never infinitely as indicated by Eq. 1.65) and multiple images blend together to form giant arcs (see fig. 1.2).

1.3.2 Thin lenses on a flat sky

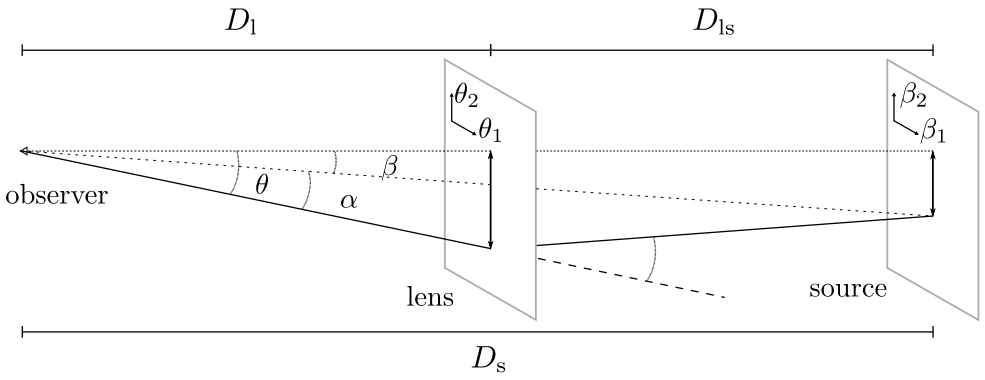

FIGURE 1.3: Illustration of gravitational lensing in the case of a single thin lens on a flat sky; angles are defined according to Eq. 1.55;

While cosmological wide field weak lensing surveys cover significant portions of the observers' sky and are concerned with the cumulative light deflection by structures over a large range in redshift, thus requiring a proper treatment of the celestial sphere and structures extended along the line of sight, lensing by galaxy clusters is well approximated by the use of flat Cartesian coordinates on the sky and by the assumption that the lenses' extent along the line of sight is negligible compared to the distances between observer, lens and source. Under these conditions, the lensing potential then simplifies to

$$\psi(\boldsymbol{\theta}) = \frac{2}{c^2}\frac{D_\mathrm{ls}}{D_\mathrm{s}D_\mathrm{l}}\int \mathrm{d}\chi' \Phi\left(f_\mathrm{K}(\chi')\boldsymbol{\theta}, \chi'\right), \tag{1.66}$$

where we introduce the shorthand $(D_\mathrm{s}, D_\mathrm{l}, D_\mathrm{ls})$ for the angular diameter distances to the source, to the lens and the distance between them. Figure 1.3 schematically illustrates this configuration.
It should be noted that the lensing potential and the convergence not only satisfy a Poisson-like equation in two dimensions (Eq. 1.59) but also that the lensing potential does not depend on the line-of-sight coordinate. Therefore the two dimensional Laplacian in Eq. (1.59) can be replaced with its three-dimensional counterpart. Recalling Poisson's equation

$$\Delta\Phi = 4\pi G\rho, \tag{1.67}$$

the convergence accordingly reduces to

$$\kappa(\boldsymbol{\theta}) = \frac{4\pi G}{c^2}\frac{D_\mathrm{l}D_\mathrm{ls}}{D_\mathrm{s}}\int \mathrm{d}\chi' \rho\left(f_\mathrm{K}(\chi')\boldsymbol{\theta}, \chi'\right) \tag{1.68}$$

[18]NASA, ESA, J. Rigby (NASA Goddard Space Flight Center), K. Sharon (Kavli Institute for Cosmological Physics, University of Chicago), and M. Gladders and E. Wuyts (University of Chicago)

and we can identify the convergence with the weighted projection of the inhomogeneous density field along the line of sight. It is convenient at this point to rewrite this equation by introducing the surface mass density

$$\Sigma(\boldsymbol{\theta}) = \int \mathrm{d}\chi' \rho\left(f_{\mathrm{K}}(\chi')\boldsymbol{\theta}, \chi'\right) \tag{1.69}$$

and the critical surface mass density

$$\Sigma_{\mathrm{crit}}(z_{\mathrm{l}}, z_{\mathrm{s}}) = \frac{\mathrm{c}^2}{4\pi \mathrm{G}} \frac{D_{\mathrm{s}}}{D_{\mathrm{l}} D_{\mathrm{ls}}} \tag{1.70}$$

to arrive at

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_{\mathrm{crit}}}. \tag{1.71}$$

As the nomenclature indicates, Σ_{crit} is a characteristic value. The existence of a point P on the lens plane with convergence $\kappa > 1$ is a sufficient condition for the lens mapping to become locally non-linear and thus leads to multiple images, since this implies the existence of a point where $\det(\mathcal{A}) < 0$. Together with the odd-number theorem (Burke, 1981), which states that thin, localised and non-singular lenses produce an odd number of images out of which at least one lies at a point where $\det(\mathcal{A}) > 0$, this in turn implies the existence of multiple images for a source at a position corresponding to P (Subramanian and Cowling, 1986). The convergence is thus often used to discern between the regimes of *weak lensing* ($\kappa < 1$), where the distortion effects are only measurable by averaging over a large enough sample of sources, essentially employing the relation (1.64), and *strong lensing* ($\kappa > 1$), where the non-linearity of the lens mapping causes clearly identifiable multiple images and severe distortions.

Another feature of strong lensing is the different arrival time of photons emitted by the same event due to their different paths. This *time delay* consists of the sum of geometric (different path length) and gravitational delay (Shapiro delay) and is thus given by (here as a delay with respect to the case without lensing)

$$\Delta t(\boldsymbol{\theta}; \boldsymbol{\beta}) = \frac{D_{\mathrm{l}} D_{\mathrm{s}}}{\mathrm{c} D_{\mathrm{ls}}} (1 + z_{\mathrm{l}}) \tau(\boldsymbol{\theta}; \boldsymbol{\beta}) \tag{1.72}$$

(Blandford and Narayan, 1986), where we have defined the so-called *Fermat potential*

$$\tau(\boldsymbol{\theta}; \boldsymbol{\beta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta}). \tag{1.73}$$

Due to their cosmological dependence via the distance kernel in Eq. (1.72), time delays of strongly lensed pulsars or supernovae events have recently been used to obtain measurements of cosmological parameters, primarily the Hubble constant H_0 (Suyu et al., 2017). Note that demanding $\partial_i \tau = 0$, i.e. applying Fermat's principle, yields again the lens equation (1.55).

1.3.3 Degeneracy of the lens mapping

The original positions, shapes and sizes of lensed sources cannot be observed, therefore the mapping for a given set of observed image positions, shapes and sizes cannot be determined uniquely but at most down to a family of mappings related by transformations that leave the observables invariant. The most commonly quoted of these invariance transformations is the so-called mass-sheet-degeneracy, where the mapping is transformed according to

$$\mathcal{A} \to \lambda \mathcal{A}, \tag{1.74}$$

which leaves the reduced shear and the image positions invariant, but scales the source size by the scalar factor λ. This transformation changes the convergence to

$$\kappa \to 1 - \lambda + \lambda\kappa, \tag{1.75}$$

which, for λ close to, but smaller than 1, corresponds to adding a sheet of constant surface mass density, hence the name. More general degeneracies have been discussed in the literature (Gorenstein, Falco, and Shapiro, 1988; Schneider and Sluse, 2014) but are not subject of this work.
In general, these degeneracies are broken by imposing a prior on the surface mass distribution of the lens, ideally informed by non-lensing data on its gravitational potential. Specifically, if lensing information is combined with other data in a joint reconstruction framework as presented in this work, degeneracies of the lens mapping can be broken as well.

Chapter 2

Galaxy clusters

The following sections will introduce our objects of interest, clusters of galaxies, in more detail, starting with their physical properties and observational signatures, before describing their use as cosmological and astrophysical probes.

2.1 Properties

2.1.1 Contents

First characterized as confined[1] peaks in the galaxy number density, galaxy clusters are far more than that. A typical cluster lies within the mass range of 10^{14}-10^{15} $M_{\odot}$. About 85 % of this mass budget is taken up by dark matter of which most is situated in an extended halo. These halos are the tentative end products of the non-linear structure formation sketched in the previous chapter. Roughly 10 % of the mass is provided by the intra-cluster medium (ICM), a hot ionized plasma dominated by hydrogen, and only a few percent of the mass consist of baryonic matter in the form of stars and galaxies (Trümper and Hasinger, 2008). These are only rough numbers, and details often depend on how exactly one defines a cluster. Contrary to the name, this nowadays is not always just identifying a literal cluster of galaxies, but rather by identifying and defining clusters based on the characteristic features of the ICM in X-ray or microwave observations (see below). Identification thresholds again differ from survey to survey, so it is not certain that two studies of the same patch of sky assign the label "cluster" to the same objects.

2.1.2 Structure

The prescriptions for non-linear structure formation in the previous chapter are clearly approximative and do not describe the real processes in full detail. In order to get closer to such a detailed description, numerical simulations are unavoidable and they form the basis for a lot of our understanding of cosmic structure today. In the context of clusters, such simulations (assuming CDM) show that they form by a succession of mergers of smaller halos, galaxies and groups of galaxies interspersed with phases of relaxation (Tormen, Bouchet, and White, 1997; Boylan-Kolchin et al., 2009; Frenk and White, 2012). Galaxy

[1] on the sky and in redshift

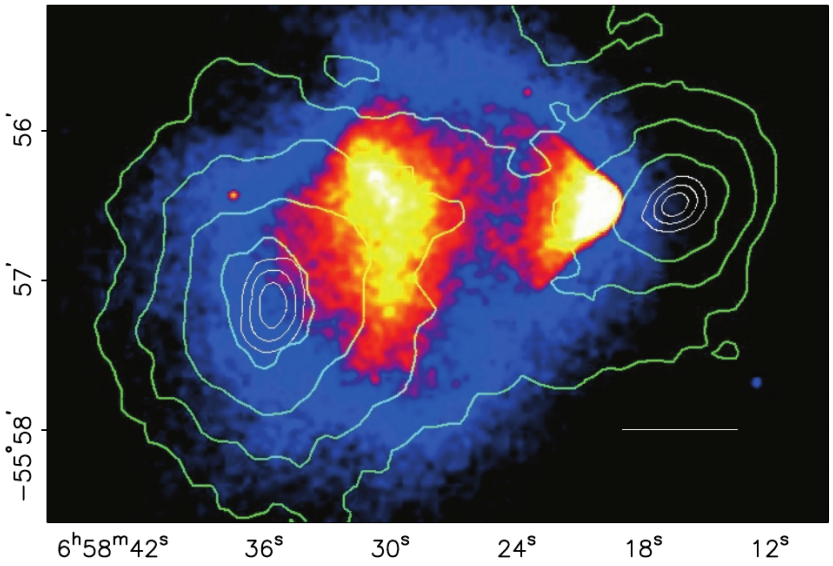

FIGURE 2.1: The Bullet Cluster (1E 0657-558), a supposedly recent merger. The heat map shows Chandra X-ray imaging of the ICM, clearly displaying shock features. Green contours depict the surface density based on weak lensing. Picture taken from (Clowe et al., 2006)

clusters vary significantly in their structure, depending on their dynamical state. While clusters that are the result of a recent merger of a number of smaller structures still show traces of that event via relatively complex morphologies across all wavelengths (see for example Figure 2.1), relaxed clusters are more symmetrical and well-behaved (e.g. see Figure 2.2).

As self-gravitating systems that seem to be dominated by collisionless dark matter, the energy budget of (at least relaxed) clusters should be governed by the virial theorem such that

$$E_{\text{total}} = E_{\text{kinetic}} + E_{\text{potential}} = E_{\text{kinetic}} - 2E_{\text{kinetic}} = -E_{\text{kinetic}}. \tag{2.1}$$

Following this any loss of total energy in the cluster, say by ejection of particles via many-body interactions, leads to an increase in kinetic energy. This means that clusters should in principle have negative heat capacity and thus no stable equilibrium state (Binney and Tremaine, 1987). Nevertheless, numerical simulations show that self-gravitating collisionless systems do indeed have long-lived metastable states in the form of dark matter halos. Embedded in these halos are the baryonic components, i.e. the ICM and galaxies, tracing the dark matter's gravitational potential.

In many applications, clusters are approximated to be spherically symmetric, such that their internal structure is characterized by a radial density profile. This is either justified by the apparent relaxed state of an individual cluster or by stacking and averaging a sample of clusters to investigate their properties in a statistical sense. It should however be noted that a cluster in general cannot

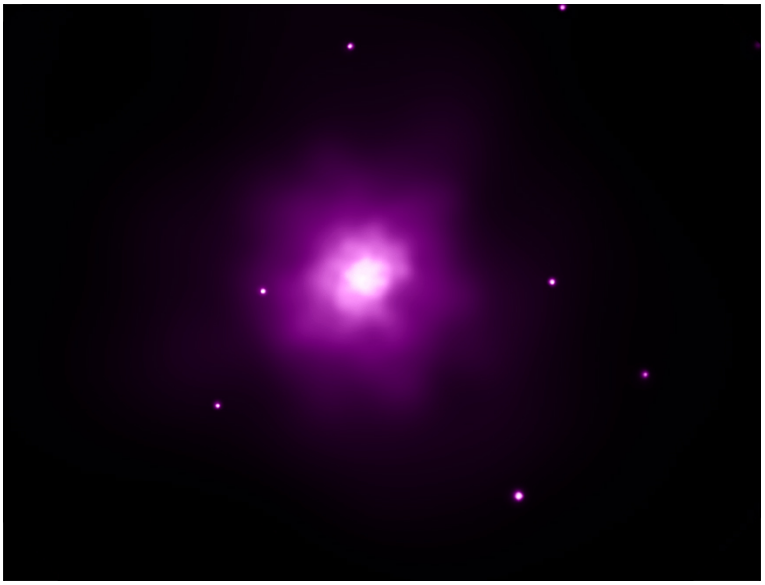

FIGURE 2.2: The galaxy cluster Abell 383 as seen by Chandra X-ray imaging of the ICM. Picture taken from http://chandra.harvard.edu

be expected to be spherical as significant triaxiality and substructure are to be expected[2] (Limousin et al., 2013). On average though, dark matter halos in numerical simulations can be reasonably well described by an universal density profile across a wide range of halo masses. One commonly used form is

$$\rho(x) = \frac{\rho_s}{x^\alpha \left(1+x\right)^{3-\alpha}}, \tag{2.2}$$

where $x = r/r_s$ is the dimensionless radius in units of the scale radius r_s, ρ_s is the amplitude at the scale radius and the hyperparameter α sets the slope of the profile. The scale radius relates to the outer radius R_{vir}, i.e. the radius within which virial equilibrium holds, via the concentration $c = R_{vir}/r_s$. For $\alpha = 1$ this profile is referred to as the Navarro, Frenk and White (NFW) profile (Navarro, Frenk, and White, 1996; Navarro, Frenk, and White, 1997), probably the most frequently used parametrisation for spherical dark matter halos, but in general values of $1 \leq \alpha \leq 1.5$ fit simulated halos well. Another parametrisation, the Einasto profile (Einasto and Haud, 1989; Merritt et al., 2006) takes the form

$$\rho(x) = \rho_s \exp\left[2n\left(x^{1/n} - 1\right)\right], \tag{2.3}$$

where in the case of dark matter halos $n \approx 5$. Both the NFW and the Einasto profile are characterised by the scale radius signifying a change in the slope and defining how concentrated a halo is. The concentration in turn is connected

[2] In short: starting from an initial Gaussian density field and employing the Zel'dovich approximation it can be shown that identical eigenvalues of the deformation tensor, corresponding to isotropic evolution, are excluded (Doroshkevich, 1970)

to the mass and redshift of a halo. Clusters of higher mass form later and as the core density reflects the background density at formation time, they are expected to have lower concentration. Several empirical or semi-analytical parametrisations of this connection exist, often of the shape

$$c(M,z) = A(M/M_{\text{pivot}})^B(1+z)^C, \tag{2.4}$$

with different pivot mass M_{pivot} and fitting parameters (A,B,C) (Seljak, 2000; Duffy et al., 2008; Macciò et al., 2008; Meneghetti et al., 2014). Such parametrisations depend on the validity of the symmetry assumptions and can be significantly biased by the fact that clusters are not spherical and contain substructure (Giocoli et al., 2014).

Numerical simulations assuming cold dark matter unanimously predict a high level of substructure for dark matter halos, increasing with host mass (Giocoli et al., 2010). Simulations that take baryonic matter into account indicate that these subhalos may be stripped of their gaseous content (Dolag et al., 2009), which likely renders them hard to observe via luminous tracers of mass.

2.1.3 Observation

Optical

As mentioned already above, historical classifications and catalogues of galaxy clusters were based on observations in the optical regime, where they appear just as overdensities of galaxies within a confined aperture and redshift range. The ACO catalogue (Abell, Corwin, and Olowin, 1989) for instance lists around 4000 such overdensities with 30 or more galaxies between redshift 0.02 and 0.2. The richness, i.e. the number of member galaxies in a cluster is a noisy estimator of the total mass (Sereno and Ettori, 2017) and can reach values of $\mathcal{O}(1000)$. The abundances of different morphological types of galaxies in clusters differs significantly from the field average. The fraction of early-type galaxies increases to up to 30 % in these dense environments and specifically giant ellipticals, like cD galaxies, are prevalent in the centers of massive galaxy clusters (Goto et al., 2003) and often the brightest cluster galaxy (BCG) is of this type. The BCG is frequently used to define the center of mass of a cluster in optical surveys.

Another observable feature that is usually measured in the optical regime are the gravitational lensing signatures caused by the clusters gravitational potential. As discussed in Section 1.3, these range from slight apparent distortions of background galaxies only measurable in a statistical sense across the field of view around the cluster, to highly magnified and warped multiple images of sources aligned with the cluster's core.

From spectroscopic measurements of the member galaxies, their velocity dispersion $\sigma_{ij}^2 = \langle v_i v_j \rangle - \langle v_i \rangle \langle v_j \rangle$ can be inferred. The dynamics of the velocity dispersion are governed by the Jeans equations (Binney and Tremaine,

1987). For instance, under the assumption of spherical symmetry and in polar coordinates, the radial Jeans equation reads

$$\frac{\partial(n\sigma_{rr}^2)}{\partial r} + \frac{n}{r}\left(2\sigma_{rr}^2 - \sigma_{\theta\theta}^2 - \sigma_{\varphi\varphi}^2\right) = -n\frac{\partial\Phi}{\partial r}, \tag{2.5}$$

where $n(r)$ is the number density of galaxies in the cluster. In principle, measurements of the velocity dispersion can therefore be used to constrain the gravitational potential and thus the mass distribution within the cluster (Capasso et al., 2018).

X-ray

Galaxy clusters are among the most luminous X-ray sources on the sky with luminosities in the range between 10^{43} and 10^{45} erg/s. The continuum radiation originates from optically thin thermal bremsstrahlung emission in the ICM, accompanied by significant line emission contributions from heavy ions, mostly highly ionized iron (Trümper and Hasinger, 2008). The emissivity of the continuum radiation is given by

$$j_{\mathrm{X}}(\omega) = \frac{16\pi^2}{3\sqrt{3}}\frac{Z^2e^6n_in_e}{m_e^2\mathrm{c}^3}g_{\mathrm{ff}}(\omega)\sqrt{\frac{2m_e}{\pi k_{\mathrm{B}}T}}\exp\left(-\frac{\hbar\omega}{k_{\mathrm{B}}T}\right), \tag{2.6}$$

where n_i and n_e are the number densities of ions and electrons, Z is the atomic number and $g_{\mathrm{ff}}(w)$ is the Gaunt factor, which in this context can safely be approximated to be 1 (Bartelmann, 2012). The frequency integrated emissivity then is

$$j_{\mathrm{X}} = \frac{16\pi^2}{3\sqrt{3}}\frac{Z^2e^6}{m_e^2\bar{m}^2\mathrm{c}^3\hbar}\sqrt{\frac{2m_e}{\pi}}\rho^2\sqrt{k_{\mathrm{B}}T}, \tag{2.7}$$

where we have replaced the number densities by the ICM density ρ and introduced the mean particle mass $\bar{m}$. Finally, the observable surface brightness of a cluster is the line-of-sight projected frequency integrated emissivity

$$S_{\mathrm{X}}(\boldsymbol{\theta}) = \int \mathrm{d}\chi j_{\mathrm{X}}(\boldsymbol{\theta},\chi). \tag{2.8}$$

Equation (2.7) shows that the X-ray signal depends quadratically on the density of the ICM and thus falls off significantly steeper with cluster radius than the latter while retaining an additional small radial dependence via the ICM temperature.

Microwave

In Section 1.2.4, the cosmic microwave background and its usefulness for cosmological inference have been introduced, but it also proves to be a valuable resource for the study of galaxy clusters. The CMB photons emitted at

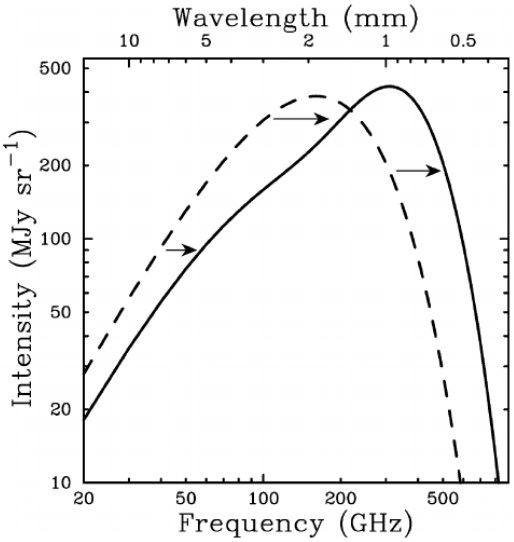

FIGURE 2.3: The distortion of the CMB black body spectrum (dashed) caused by the thermal SZ effect (solid). The effect is strongly exaggerated to improve visibility. Figure taken from Carlstrom, Holder, and Reese (2002)

the time of recombination do not travel entirely unperturbed. Besides gravitational effects like lensing and the integrated Sachs-Wolfe effect (Rees and Sciama, 1968), there are also electromagnetic interactions affecting the photons. Traversing through comparably dense and ionized gas like the ICM, about 1% of them undergo Compton scattering by thermal electrons. Due to the very high temperature of the ICM plasma, this typically leads to an energy boost for the photon by roughly $k_B T_e / m_e c^2$ (Carlstrom, Holder, and Reese, 2002)(m_e and T_e are rest mass and temperature of the ICM electrons) and thus a slight distortion in the otherwise almost perfect black body spectrum of the CMB photons, as depicted in Figure 2.3. This inverse Compton scattering is usually referred to as the thermal Sunyaev-Zel'dovich (tSZ) effect (Sunyaev and Zeldovich, 1980). It alters the specific intensity of the CMB with respect to the Planck spectrum B_ω as

$$\frac{\Delta I_{\mathrm{tSZ}}(x, \boldsymbol{\theta})}{B_\omega} = g(x) y(\boldsymbol{\theta}), \tag{2.9}$$

as a function of position on the sky $\boldsymbol{\theta}$ and dimensionless frequency $x = \hbar\omega / k_B T_{\mathrm{CMB}}$. The frequency dependence is given by

$$g(x) = \frac{x^4 e^x}{(e^x - 1)^2} \left(\frac{e^x + 1}{e^x - 1} - 4 \right) \tag{2.10}$$

and depicted in Figure 2.4. Photons with frequencies below the zero-crossing value of $x \approx 3.83$, corresponding to 217 GHz, get up-scattered to frequencies above this threshold. This characteristic signature makes the thermal SZ

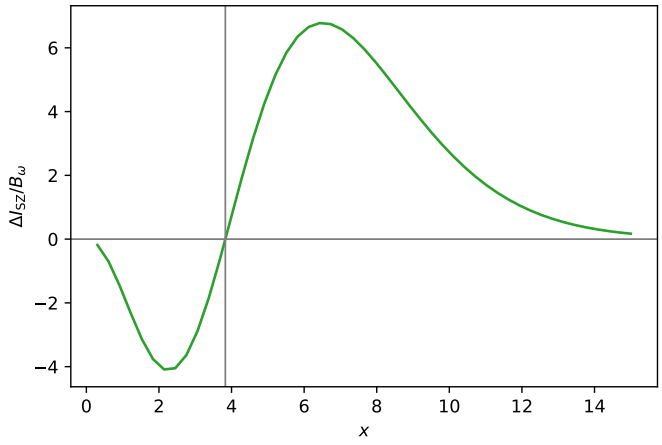

FIGURE 2.4: The frequency dependence of the thermal SZ effect, where $x = \hbar\omega/k_B T_{CMB}$.

effect easily identifiable. The amplitude of the tSZ effect is quantified by the Compton-y parameter

$$y(\boldsymbol{\theta}) = \int \mathrm{d}\chi \frac{k_B T_e}{m_e c^2} \sigma_T n_e(\boldsymbol{\theta}, \chi). \tag{2.11}$$

It is the line-of-sight projection of an effective pressure given by the product of the mean energy change and the scattering probability $\sigma_T n_e(\boldsymbol{\theta}, \chi)$, where σ_T is the Thomson cross-section and $n_e(\boldsymbol{\theta}, \chi)$ the number density of electrons. The signal strength is independent of redshift, which makes the tSZ effect well suited for the detection clusters at high redshift, even though in practise detectability is still slightly dependent on redshift via the angular size of a cluster on the sky and thus the angular diameter distance.

2.2 Clusters as cosmological probes

Here I very briefly list some of the ways galaxy clusters can be used for cosmological inferences. This is not an exhaustive collection and I refer the reader to reviews like Allen, Evrard, and Mantz (2011) for further details.

2.2.1 Mass

In Section 1.2.5 we introduced the Press Schechter mass function and how it relates to cosmological parameters. Similar cosmological dependencies hold for more sophisticated mass functions and those calibrated to numerical simulations. As galaxy clusters reside in the most massive dark matter halos, they probe the high-mass exponential cut-off of the mass function and thus their abundance is a very sensitive probes of the underlying cosmology (Allen, Evrard, and Mantz, 2011; Kravtsov and Borgani, 2012; Hagstotz et al., 2018).

There is however no single preferred definition of mass that is the same for analytical or numerical predictions on the one side and observations on the other. Halos in simulations are usually identified and assigned mass either via friends-of-friends algorithms (Davis et al., 1985) or spherical overdensities (Springel et al., 2001). The latter definition is closer related to analytical halos based on spherical collapse (see Section 1.2.5) but the choice of overdensity threshold is a matter of debate (Coe, 2010). Often an overdensity of of $200\rho_{\mathrm{crit}}$ is used to approximate the virial overdensity of $178\rho_{\mathrm{crit}}$ in a matter-dominated universe. On the observational side, masses are usually assigned by measuring aperture averaged values of mass proxies like richness, velocity dispersion or X-ray luminosity and connecting them to a mass definition of choice via analytically, numerically or empirically motivated scaling relations (Sereno and Ettori, 2017).

Just as the masses of galaxy clusters, also the abundance and masses of subhalos within a cluster carry information on dark matter and structure formation (Jauzac et al., 2016). Deviations from the expected abundances and distribution might indicate flaws in the set of assumptions for CDM but conclusive claims are often hindered by inconsistent mass definitions and substructure identification (Schwinn et al., 2018).

2.2.2 Clustering

Numerical simulations as well as observations show that non-linear structure formation transforms the initially Gaussian density field into a web-like large scale structure made of sheets and filaments of overdensities surrounding voids depleted of matter. Galaxy clusters represent the nodes of this cosmic web and as every identifiable structure in this web is a more or less biased tracer of the underlying density field (Pollina et al., 2017), clusters are no exception. Their clustering, i.e. the 2-point correlation function of cluster positions can be used to constrain the matter power spectrum and thus cosmology (Schuecker et al., 2003; Hütsi and Lahav, 2008; Mana et al., 2013).

2.2.3 Lensing

The core regions of massive clusters often act as strong gravitational lenses with magnifications that easily reach factors of $\mathcal{O}(10)$ but also much higher magnifications around 200 are known (Zitrin and Broadhurst, 2009) and in at least one case a star behind a cluster has been magnified by a factor of roughly 2000 (Kelly et al., 2018). This shows that strong lensing clusters can be used to study distant objects that would otherwise be excluded from observations due to resolution constraints or because they would be too faint. This use of clusters as "cosmic telescopes" has allowed to discover and study extremely distant and thus early galaxies up to redshifts $z \approx 7$ (Kneib et al., 2004).

Another cosmological use for strong lensing clusters are measurements of the time delay between strongly lensed images of time dependent sources like supernovae or quasars. As already stated in Section 1.3.2, measurements of

this time delay allow inference on the Hubble constant H_0 via the time delay distance (Suyu et al., 2017).

The caveat to both, investigations of high redshift galaxies and inferences of the Hubble constant, is that in most cases knowledge about the mass configuration of the lens and its surroundings is required.

2.3 Clusters as astrophysical laboratories

ICM physics

The formation of a galaxy cluster involves complex hydrodynamical processes as the ICM collapses alongside the dark matter. Infall of gas at the outskirts of collapsing halos can happen at supersonic speed, which leads to the formation of pronounced shock fronts. Apart from quasi-isotropic infall, simulations as well as observations suggest that gas streams towards the center of mass along filamentary structures. At least in relaxed clusters, the gas is often assumed to be in hydrostatic equilibrium (HE), with only thermal pressure compensating the gravitational potential, but indications for deviations from HE, especially in the outer regions of clusters and in recent mergers are prevalent, again in simulations (Shi and Komatsu, 2014; Biffi et al., 2016) and observations (Hurier and Angulo, 2018).

Scaling relations between ICM observables like the X-ray luminosity or the Compton-y parameter and mass rely, beside hydrostatic equilibrium, on the self-similarity of clusters, i.e. it is assumed that clusters of different size and mass are indeed just scaled versions of one another. Observations do however show that self-similarity needs to be broken in the central regions of the cluster (Voit, 2005; Allen, Evrard, and Mantz, 2011 and references therein). Several mechanisms like radiative cooling (Voit and Bryan, 2001), preheating (Evrard and Henry, 1991) and heating by AGN feedback (Forman et al., 2005; McNamara and Nulsen, 2007) as well as their interplay may play a role in this phenomenon and detailed observations in X-rays and microwaves offer a way to test these contributions.

The metallicity of the ICM allows insight on the chemical enrichment of the Universe, on star formation history and on the yield of several types of supernovae (Böhringer and Werner, 2010).

Chapter 3

Reconstruction methods

The goal of this work is to provide a method to infer the (projected) matter distribution of a galaxy cluster from a combination of observational constraints. The following chapter will discuss general properties of such reconstruction methods as well as the specific methods chosen here. As this project was and is a collaborative and long-running effort, much of what is described here builds on published work of collaborators and predecessors (Bartelmann et al., 1996; Merten et al., 2009; Konrad et al., 2013; Sarli et al., 2014; Majer et al., 2016; Merten, 2016; Tchernin et al., 2018), as well as myself (Huber et al., 2018).

3.1 Classifying methods

The outcome of this work is a mesh-free free-form linear-Gaussian maximum posterior estimator of the projected gravitational potential by direct inversion using joint data. Since this is a rather opaque sentence, the following section will attempt to unpack the previous terms and put them into context by providing a short guide to the various types and categories in the business of inference. Afterwards, we will revisit the unwieldy sentence above and describe the methods chosen for this work in section 3.2.

3.1.1 Types of modelling

Models used in reconstructions usually fall in one of the following categories.

Parametric methods are based on analytical models with comparably few parameters (the number of parameters P is much smaller than the number of data points / constraints N) to describe a given data set. Such models draw their motivation either from theoretical predictions or are based on empirical knowledge alone. The former case, particularly in the field of Astrophysics and Cosmology, explicitly includes predictions based on numerical simulations of the system in question. By construction, the number of data points per parameter to be fitted is large, allowing (though not guaranteeing) strong constraints and thus high precision, i.e. a low statistical error. The downside is that such reconstructions are in danger of being significantly biased, i.e. suffer from increased systematic error (low accuracy) in cases where the parametric model turns out to be an inadequate description of the physical system studied.

Whether this is the case is not necessarily detectable by standard goodness-of-fit measures and subsequently may lead to wrong inferences based on the obtained reconstruction. Therefore parametric modelling is most successful when the assumptions adopted by the model are fulfilled to a high degree, either by an individual system or on average by a sample of systems. Additionally, since the dimensionality of their parameter space is often rather low, they are ideally suited for use of inference methods that explore that parameter space to at least some extend like for example Monte-Carlo Markov-chain samplers.

Local **free-form** methods on the other hand are characterized by making little assumptions on a specific model and rather describe the system by a large number of parameters[1] ($P \lesssim N$), essentially directly fitting the target variable at P points in the respective domain. While the precision of course suffers from this and the statistical error is increased, these methods are intended to be less prone to systematic errors and biasing and thus to more accurately describe the reconstructed system. They are however not free of prior assumptions. As the number of parameters increases the danger of overfitting – that is mistakenly modelling noise patterns as part of the signal – rises. Therefore, free-form reconstructions require some sort of regularisation which takes the form of an additional prior on the presumed shape or characteristics of the model (e.g. smoothness, etc.). Free-form methods are best employed in situations where the data quality is exquisite but where it is uncertain that a specific parametric model is applicable. These conditions are met especially for the most massive clusters of galaxies, which, as discussed above, have strong signals in lensing, X-ray and tSZ observations and contain a large number of member galaxies tracing their Newtonian potential but on the other hand are not necessarily relaxed and well described by any simple parametric model. Another common use case are situations where a theoretically or otherwise motivated model is just not available yet.

Non-local (decomposition) methods fall somewhat in-between. They do not presume any specific functional shape, so in that sense they are free-form methods. But rather than fitting a large number of parameters at different points in the domain of the model, they rely on a decomposition of the data in some non-local basis set and then fit a – not necessarily large – number of coefficients of that decomposition. The classical example are Fourier analyses, where the signal is Fourier transformed and described by a limited set of spectral coefficients. But also shapelet and wavelet decompositions are abundantly used (Daubechies, 1990; Refregier, 2003). Based on these decomposition coefficients, further transformations to a target variable, i.e. the underlying physical quantity, can be performed. The advantage of these methods lies in the availability of fast high-performance methods of decomposition, the often easier means of noise reduction and the high flexibility to describe very different and apparently complex data sets with comparably few coefficients. The downside is that the majority of methods in this category are inherently non-local, thus correlating intrinsically uncorrelated data points, and often

[1] which makes the custom of referring to them as non-parametric so counter-intuitive

necessitate ad-hoc boundary treatment when dealing with finite data (e.g. zero-padding to enforce periodic boundary conditions in Fourier methods).

3.1.2 Types of support structure

Especially in local free-form methods the primary choice is where to fit the data, that is on how many and which points in the data domain should the target variable be reconstructed. In Astronomy a lot of data come in the form of discrete two-dimensional maps or images of a part of the sky, the field of view (or footprint, when speaking of large surveys). The question then translates to: How to best sample the field of view with support nodes for the reconstruction?

In this context, a **mesh** is a collection of *connected* nodes, with the connection adhering to predefined geometry (Liu, 2002). The most common examples are regular rectangular grids, but also Delaunay triangulations (the dual of a Voronoi graph) fall in this category. Values of the field that is to be discretized by the mesh are assigned to nodes (Delaunay triangulation) or the volume elements of the mesh (pixels or voxels, depending in dimensionality). This discretization with an a priori known relation between neighbouring nodes allows to approximate (partial) differential equations by algebraic ones in a relatively straight-forward manner, thus simplifying the task of solving them. This forms the basis of traditional approaches to finite difference (FD), finite element (FE) and finite volume methods (FV).
Structured grids are meshes which follow a global coordinate structure and in which the neighbouring elements for each cell are easily identifiable. Given the basic properties of the mesh (dimensionality, size, resolution, coordinate structure) the geometric properties, i.e. cell coordinates, cell size and neighbouring cells can be determined from the cell index alone. Regular grids are structured grids with cells identical in shape and size. A simple Cartesian grid is the prime example for regular grids.
Unstructured grids still adhere to certain constraints on their geometry but adjacent cells are not as easily identified and the size and aspect ratios of cells may vary. An example is the aforementioned Delaunay triangulation that tessellates the plane with triangles in a unique way, namely such that the minimum opening angle is maximized (Press et al., 1992). Figure 3.1 illustrates these different types of meshes.

In contrast, **mesh-free** support is given by a set of nodes *without* any predefined connection or tessellation/sampling pattern. These nodes do not discretize the domain but rather represent it, as there is no defined graph connecting the nodes and thus dividing up the domain. This ambiguity makes FD methods somewhat more complicated but also offers some substantial advantages. Regular grids can deal with situations that require different resolutions depending on the position in the domain via adaptive mesh refinement (AMR) methods (Berger and Oliger, 1984), effectively replacing the cells in the respective area with a finer subgrid. This however substantially complicates the usual FD or FE methods and requires careful treatment of the boundary

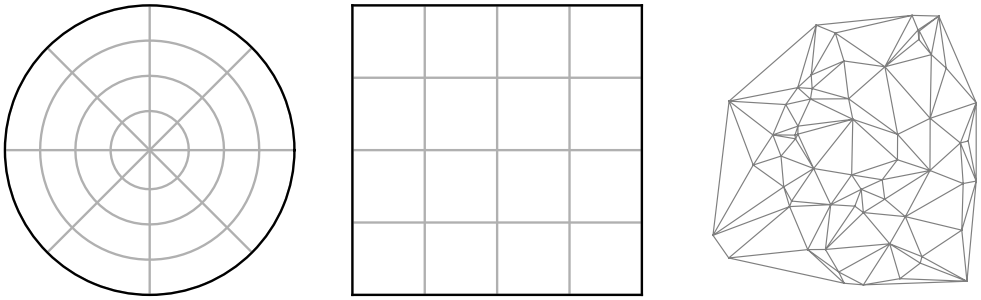

FIGURE 3.1: Examples of meshes: (*left*) a structured polar grid, (*middle*) a regular Cartesian grid, (*right*) an unstructured Delaunay triangulation;

between grid and subgrid. And even unstructured grids have to be recomputed entirely whenever individual nodes are added, removed or even just shifted in position, which can be computationally costly. Mesh-free methods do not suffer from these problems at all. The resolution can easily be changed by just adding more nodes where needed, without almost any computational overhead since no global mesh-structure has to be recomputed. Section 3.2 will demonstrate in detail how the interpolation and differentiation operations needed in this work can be performed on a mesh-free support domain. Other examples for mesh-free algorithms include smoothed particle hydrodynamical simulations and some classical N-body codes (Gingold and Monaghan, 1977; Lucy, 1977; Liu, Liu, and Li, 2004; Trenti and Hut, 2008).

3.1.3 Types of regression

A fundamental equation for conditional probabilities and therefore the statistical inference[2] of a set of model parameters q given a set of data points d is Bayes' theorem (Bayes, 1763)

$$p(q|d,\alpha) = \frac{p(d|q)p(q|\alpha)}{p(d|\alpha)}, \tag{3.1}$$

where the *posterior* $p(q|d)$ is the probability of q given d, the *likelihood* $p(d|q)$ is the probability of the data given the model and the *prior* $p(q|\alpha)$ quantifies prior knowledge (or assumptions) on the model and may itself depend on a set of *hyperparameters* α. Lastly the *evidence* or marginal likelihood $p(d|\alpha) = \int \mathrm{d}q p(d|q)p(q|\alpha)$ can (at least here) be seen as a normalization. With Bayes' theorem in mind, two ways of regression can be distinguished by the choice of objective functional that is to be optimized.

Maximum likelihood methods try to maximize the probability of the observed data, given the model by optimizing the model parameters. This is the simplest approach, ignoring any prior assumptions. Usually the optimization actually is done by minimizing the (negative) log-likelihood, which in the

[2]This work is composed with a moderately Bayesian mindset

case of Gaussian likelihoods without covariance immediately reduces to the familiar least-squares fit.

A bit more Bayesian, **maximum posterior** methods take prior assumptions and knowledge into account and maximize the probability of the model parameters given the measured data, usually by minimizing the (negative) log-posterior. Note that regularized maximum likelihood methods fall into this category as well, since regularization terms, usually added to the log-likelihood to ensure certain properties of the solution (e.g. smoothness) are nothing but quantified prior knowledge on the model, i.e. a prior. This then naturally also includes **maximum entropy** methods which in this context are just regularized maximum likelihood methods where the regularization term takes the functional form of a (relative) Shannon entropy (Shannon, 1948; Press et al., 1992).

A related distinction is that of forward versus backward modelling. In our context this boils down to the way the log-likelihood is written. If the minimization is performed in the data space, i.e. the difference of data points and predictions based on the model

$$\log p(d|q) \propto (d - m(q)) \tag{3.2}$$

is minimized, one speaks of forward modelling. Conversely, backward modelling describes case when the model[3] m is inverted to transform the data to the parameter space and the minimization is performed there, i.e.

$$\log p(d|q) \propto \left(m^{-1}(d) - q\right). \tag{3.3}$$

As for the actual optimization, methods typically fall in the categories of direct inversion and sampling techniques. While the former are predominantly used for models that are at least approximately linear in the parameters and work by inverting a system of equations describing the minimization of the objective functional (log-likelihood or log-posterior) with respect to the parameters, the latter are used for more complicated models and work by sampling the parameter space and evaluating the objective functional for each point sampled this way to approximate the optimum. Typical representatives of sampling techniques are Monte Carlo Markov chain (MCMC) algorithms and nested sampling techniques (Hastings, 1970; Press et al., 1992; Skilling, 2004).

3.2 A joint reconstruction framework

As stated above, the method described in this work enables mesh-free free-form reconstructions of the lensing potential ψ of a galaxy cluster. Mesh-free support was chosen, because we want to combine different probes of ψ in a joint fit and the spatial scales of these probes differ significantly. While for example weak lensing measurements typically span several (tens of) arcminutes

[3] sometimes called response

with about $\mathcal{O}(20)$ data points per arcmin^2 but are increasingly hard to obtain for the inner core of the cluster as member galaxies get in the way, strong lensing features like multiple images are usually confined to that innermost region of the field of view with typical distances between data points of arcseconds. The intrinsic adaptability of mesh-free methods allows for a straight-forward combination of such different probes. Therefore a set of reconstruction nodes is defined and the value of ψ is fitted for each of these nodes, marking the free-form character of the method.

To do so, we assume that all measurements of the different probes (shear, multiple images, X-rays, etc.) are taken independently and that each probe has approximately Gaussian errors. Thus we are able to define a predictive posterior distribution for ψ as

$$p(\boldsymbol{\psi}|\boldsymbol{d},\alpha) \propto p(\boldsymbol{d}|\boldsymbol{\psi})p(\boldsymbol{\psi}|\alpha) = \prod_i p_i(\boldsymbol{d}_i|\boldsymbol{\psi}) \cdot p(\boldsymbol{\psi}|\alpha), \tag{3.4}$$

where the joint likelihood $p(\boldsymbol{d}|\boldsymbol{\psi})$ of all probes can be expressed as the product of the individual likelihoods due to their independent measurements. With this and the assumption that these likelihoods are approximately Gaussian, we can express the negative log-posterior as a sum of χ^2 terms

$$\chi^2_{\text{joint}}(\boldsymbol{\psi}|\boldsymbol{d}) = \sum_i \chi^2_i(\boldsymbol{\psi}|\boldsymbol{d}_i) + \chi^2_{\text{reg}}(\boldsymbol{\psi}|\alpha), \tag{3.5}$$

where we have called the negative log-prior χ^2_{reg}, as it will act as a regularization term (see below). The third condition we impose on our modelling technique is that the log-likelihoods are approximately quadratic in ψ. If this is fulfilled we can find the maximum posterior estimate of ψ at any node k by first minimizing χ^2_{joint} with respect to ψ_k

$$\frac{\partial \chi^2_{\text{joint}}}{\partial \psi_k} \stackrel{!}{=} 0, \tag{3.6}$$

and then solving the subsequent linear system of equations (LSE)

$$B_{kl}\psi_l = V_k \tag{3.7}$$

for ψ.

A word on notation: in this section and the following ones, vector notation, i.e. boldface typesetting, can either describe two dimensional vectors on the sky, as for example in the case of coordinates ($\boldsymbol{\theta}$, $\boldsymbol{\beta}$, $\boldsymbol{x}$), or it describes an at least N-dimensional data or parameter vector, where N is the number of reconstruction nodes in the mesh-free domain. The latter is for example the case in Eq. (3.5), where the $\boldsymbol{d}_i$ are data vectors over the domain and $\boldsymbol{\psi}$ is the vector of ψ values on all nodes. This typographic degeneracy is unfortunate, but context should make the distinction obvious and where it does not, the dimensionality and nature of the respective vector will be stated explicitly.

As the definition of the lensing potential of a thin lens in Eq. (1.66) involves both the distance to the lens and to the source, a further minor complication arises. As the different probes of ψ are based on sources at in parts vastly different redshifts these constraints have to be scaled to a common redshift in order to consistently reconstruct the lensing potential throughout the field of view and using all available data sets. To do so, the cosmic weight function

$$Z(z_s) \equiv \frac{\lim_{z\to\infty} \Sigma_{\text{crit}}(z_l, z)}{\Sigma_{\text{crit}}(z_l, z_s)} \Theta(z_s - z_l) = \frac{D_{ls} D_\infty}{D_{l\infty} D_s} \Theta(z_s - z_l) \tag{3.8}$$

is introduced (Bartelmann and Schneider, 2001) and applied to all data sets used, which by construction are approximately linear in ψ. This way, all sources are scaled to a formally infinite redshift, although a finite but large reference value is used in practice, and the reconstruction is performed for this scaled scenario. The Heaviside step function $\Theta(z_s - z_l)$ rules out lensing of sources in front of the lens. Also all estimates of ψ based on non-lensing constraints that enter the reconstruction as introduced in section 3.5 will be scaled in this fashion to allow a consistent treatment.

Radial basis functions for FD and interpolation

As can be expected from section 1.3 and will be detailed below, the projected potential ψ is related to several of the observables via angular derivatives, which means that a FD implementation is required in order for Eq. (3.7) to describe a system of algebraic equations that can be solved via standard linear algebra methods. Apart from finite differences, the implementation described in section 3.6 will require the ability to interpolate fields from one mesh-free domain to another. Both can be achieved via the use of radial basis functions (RBF). This means any function $\phi(\|\boldsymbol{x} - \boldsymbol{x}_0\|)$[4] which only relies on the distance of an evaluation point $\boldsymbol{x}$ to a specified reference point $\boldsymbol{x}_0$. A few typical choices are collected in Table 3.1. This section will briefly present how RBFs can be used to perform interpolations and differentiations on mesh-free domains, which is mostly based on Merten (2016). For more information on this see for example Fornberg, Lehto, and Powell (2013) and Flyer et al. (2016).

By centring an RBF on each of the support nodes of a mesh-free domain, the **Interpolation** $\tilde{f}$ of a function f defined on each of those nodes can be expressed as

$$\tilde{f}(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i \phi(\|\boldsymbol{x} - \boldsymbol{x}_i\|), \tag{3.9}$$

with the interpolation weights λ_i and where n is the number of neighbouring nodes in consideration, which can typically be chosen much smaller than the total number of support nodes in the domain. Since the interpolation is

[4] $\|\cdot\|$ here denotes the Euclidean L_2 norm

Name	Form $\phi(r)$
Gaussian	$\exp\left(-(\varepsilon r)^2\right)$
multiquadratic	$\sqrt{1+(\varepsilon r)^2}$
inverse multiquadratic	$1/\sqrt{1+(\varepsilon r)^2}$
inverse quadric	$1/\left(1+(\varepsilon r)^2\right)$
Polyharmonic splines	r^{2m-1}, $m \in \mathbb{N}$
	$r^{2m}\log r$

TABLE 3.1: some typical choices of RBFs; ε is the shape (hyper-) parameter;

supposed to exactly recover the original function on each of the support nodes

$$\tilde{f}(\boldsymbol{x}_i) = f(\boldsymbol{x}_i) \equiv f_i, \tag{3.10}$$

we end up with a linear system of equations

$$\underbrace{\begin{pmatrix} \phi(\|\boldsymbol{x}_1-\boldsymbol{x}_1\|) & \cdots & \phi(\|\boldsymbol{x}_1-\boldsymbol{x}_n\|) \\ \vdots & \ddots & \vdots \\ \phi(\|\boldsymbol{x}_n-\boldsymbol{x}_1\|) & \cdots & \phi(\|\boldsymbol{x}_n-\boldsymbol{x}_n\|) \end{pmatrix}}_{\mathcal{F}} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}, \tag{3.11}$$

that can be solved for the weights λ_i. The resulting accuracy of this procedure (as well as of the FD scheme described below) depends on the specific choice of radial basis function. In the case of Gaussian RBFs, which will be adopted throughout this work, the shape parameter ε remains free and can significantly affect the quality of interpolation. It can however be optimized beforehand by testing it on an analytic test function with similar scaling behaviour as the field that is supposed to be reconstructed (Fornberg, Lehto, and Powell, 2013; Merten, 2016).

Finite differencing in general approximates a derivative (or really any linear operator) Df of a function at a point $\boldsymbol{x} = \boldsymbol{x}_\mathrm{c}$ by a weighted sum over the n neighbouring points

$$Df(\boldsymbol{x}) \approx \sum_{i=1}^{n} w_i f(\boldsymbol{x}_i). \tag{3.12}$$

The weights w_i that form the FD stencil can be obtained in various ways (for a didactic introduction to some see for example Fornberg (1998)), here we opt for the following. The FD approximation is demanded to exactly recover

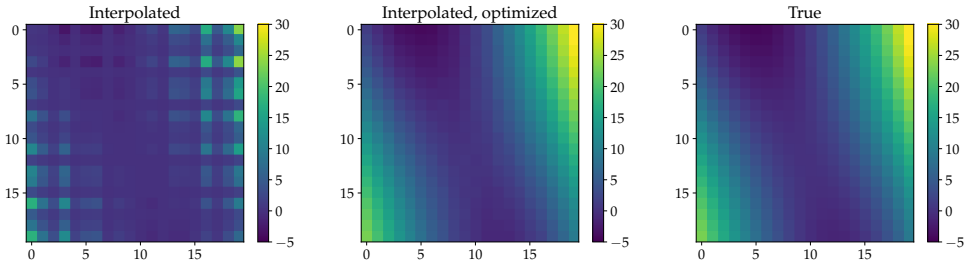

FIGURE 3.2: Illustration of a Gaussian RBF-based interpolation of a polynomial function from a 10 by 10 regular grid to a 20 by 20 grid. (*left*) interpolation with un-optimized shape parameter, (*middle*) with shape parameter optimized using a trigonometric function, (*right*) the analytic polynomial function on the finer grid;

the analytically known derivatives of the RBF interpolation above, which, in analogy to Eq. (3.11) leads to the linear system of equations

$$\mathcal{F}\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} D\phi(\|\boldsymbol{x}-\boldsymbol{x}_1\|) \\ \vdots \\ D\phi(\|\boldsymbol{x}-\boldsymbol{x}_n\|) \end{pmatrix}. \tag{3.13}$$

Note that, if D was the identity operator, Eq. (3.11) and therefore the interpolation (the zeroth FD derivative) is recovered. The FD stencil obtained by solving Eq. (3.13) approximates the derivative Df only at the point $\boldsymbol{x}_c$, which distinguishes this approach from FD stencils on regular grids, where (except for boundaries) the same stencil can be applied all over the grid. Consequently the weights have to be computed for the $n-1$ nearest neighbours of each node in the domain individually.

Following Flyer et al. (2016) (and references therein) the accuracy of this FD and interpolation method can be improved by additionally demanding that monomial test functions to a certain (low) order are recovered by it exactly as well. In the example case of monomials up to first order, this changes Eq. (3.13) to

$$\begin{pmatrix} & & & 1 & x_{1,1} & x_{2,1} \\ & \mathcal{F} & & \vdots & \vdots & \vdots \\ & & & 1 & x_{1,n} & x_{2,n} \\ 1 & \cdots & 1 & & & \\ x_{1,1} & \cdots & x_{1,n} & & 0 & \\ x_{2,1} & \cdots & x_{2,n} & & & \end{pmatrix}\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = \begin{pmatrix} D\phi(\|\boldsymbol{x}-\boldsymbol{x}_1\|) \\ \vdots \\ D\phi(\|\boldsymbol{x}-\boldsymbol{x}_n\|) \\ D1 \\ Dx_{1,c} \\ Dx_{2,c} \end{pmatrix}. \tag{3.14}$$

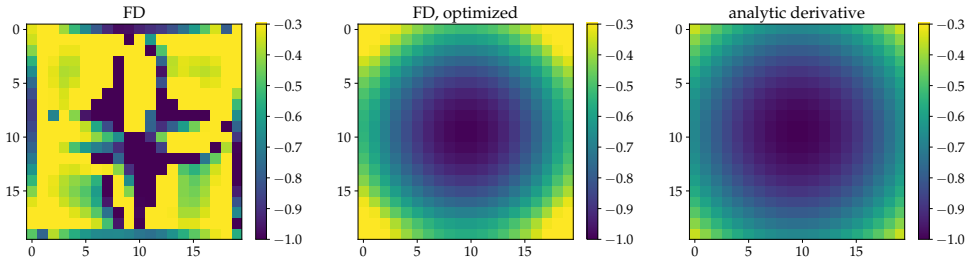

FIGURE 3.3: Illustration of a Gaussian RBF-based FD 2nd derivative ($\partial^2/\partial x \partial y$) of a polynomial function on a 20 by 20 regular grid. (*left*) FD with un-optimized shape parameter, (*middle*) with shape parameter optimized using a trigonometric function, (*right*) the analytic derivative of the polynomial;

The ν_i introduced here are not part of the FD stencil and can be viewed as Lagrange multipliers of a constrained optimization procedure. With this method, FD versions of all derivatives and combinations of derivatives needed in this work can be defined. Specifically we define the linear operators

$$\mathcal{D}_i \approx \frac{\partial}{\partial \theta_i} \tag{3.15}$$

$$\mathcal{K} \approx \frac{1}{2}\left(\frac{\partial^2}{\partial \theta_1^2} + \frac{\partial^1}{\partial \theta_2^2}\right) \tag{3.16}$$

$$\mathcal{G}_1 \approx \frac{1}{2}\left(\frac{\partial^2}{\partial \theta_1^2} - \frac{\partial^1}{\partial \theta_2^2}\right) \tag{3.17}$$

$$\mathcal{G}_2 \approx \frac{\partial^2}{\partial \theta_1 \partial \theta_2}, \tag{3.18}$$

which, when applied to the lensing potential ψ, recover the two components of the deflection angle, the convergence and the two components of the complex shear respectively (see section 1.3.1).

3.3 Lensing reconstructions

3.3.1 Weak lensing

As mentioned in section 1.3, gravitational lensing to linear order distorts size and shape of background sources, i.e. galaxies behind the lens. Equation (1.64) in particular shows how an ellipticity induced by lensing of a circular source is related to the reduced shear g, the complex-valued representation of the anisotropic part of the lensing Jacobian. Measuring the ellipticity of a lensed background galaxy can constrain the gravitational potential and therefore the mass distribution of a gravitational lens, which in this work will be a cluster of galaxies.

The immediate observable in weak lensing studies is a noisy measurement of the lensed galaxies' surface brightness $I(\boldsymbol{\theta})$. First defining the center of brightness

$$\bar{\boldsymbol{\theta}} \equiv \frac{\int \mathrm{d}^2\theta I(\boldsymbol{\theta})\boldsymbol{\theta}}{\int \mathrm{d}^2\theta I(\boldsymbol{\theta})} \tag{3.19}$$

as a reference coordinate and calculating the second moment tensor of the surface brightness distribution

$$Q_{ij} = \frac{\int \mathrm{d}^2\theta I(\boldsymbol{\theta})(\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int \mathrm{d}^2\theta I(\boldsymbol{\theta})}, \tag{3.20}$$

we can define the complex ellipticity of that distribution as

$$\epsilon = \frac{Q_{11} - Q_{22} + 2\mathrm{i}Q_{12}}{Q_{11} + Q_{22} + 2\left(Q_{11}Q_{22} - Q_{12}^2\right)^{1/2}}. \tag{3.21}$$

Note that there are several other ways to define an ellipticity based on the second moment tensor, but the one given here turns out to be the most convenient for the purposes of this work. The second moment tensor of the observed surface brightness is related to that of the intrinsic (unlensed) brightness of the source via $\mathcal{Q}\mathrm{s} = \mathcal{A}(\bar{\boldsymbol{\theta}})\mathcal{Q}\mathcal{A}(\bar{\boldsymbol{\theta}})^{\mathrm{T}}$, and the observed ellipticity as a function of the intrinsic source ellipticity ϵ_s is

$$\epsilon = \begin{cases} \dfrac{\epsilon_\mathrm{s} + g}{1 + g^\star \epsilon_\mathrm{s}} & \text{if } |g| \leq 1 \\ \\ \dfrac{1 + g\epsilon_\mathrm{s}^\star}{\epsilon_\mathrm{s}^\star + g^\star} & \text{else.} \end{cases} \tag{3.22}$$

A star denotes complex conjugation. Under the assumption that the intrinsic alignment of galaxies, that are close by each other in projection but distant in redshift, is negligible[5], the orientations of these galaxies are uncorrelated and their intrinsic ellipticity on average vanishes

$$\langle \epsilon_\mathrm{s} \rangle = 0. \tag{3.23}$$

This leads to the simplified relation

$$\langle \epsilon \rangle = \begin{cases} g & \text{if } |g| \leq 1 \\ \\ \dfrac{1}{g^\star} & \text{else.} \end{cases} \tag{3.24}$$

[5]This assumption is far from guaranteed to be valid and significant effort is made to reliably quantify the bias introduced this way (see for example Tugendhat and Schäfer, 2018 and references therein)

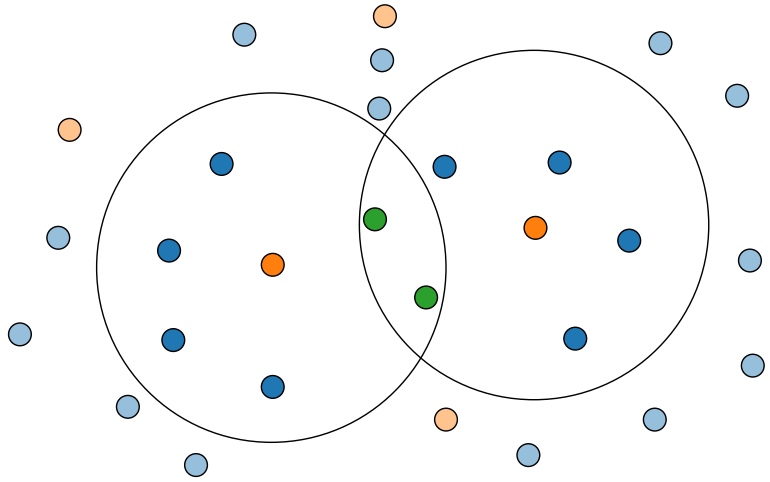

FIGURE 3.4: The averaging scheme for shear measurements on a mesh-free domain. Reconstruction nodes are depicted in orange, galaxies with ellipticity measurements in blue and those galaxies that contribute to the average ellipticity of two nodes in green.

Restricted to the weak lensing regime, where $|g| \leq 1$ can safely be assumed, Eq. (3.24) defines how ellipticity measurements are used in this work to constrain the gravitational potential of the lens. Given a catalogue of ellipticity measurements for galaxies behind the lens and after choosing an appropriate set of reconstruction nodes sufficiently sampling the respective field of view (see section 3.6), ellipticities of galaxies in the vicinity of each node are averaged over to get an estimate of $\langle \epsilon \rangle(\boldsymbol{\theta})$. This averaging process introduces covariances between neighbouring reconstruction nodes, as they might "share" galaxies. This is shown in Figure 3.4. The resulting covariance matrix is given by

$$C_{ij} = \frac{2n_{ij}}{(n_i + n_j)} \sigma_i \sigma_j, \tag{3.25}$$

where $n_{i/j}$ and $\sigma_{i/j}$ are the total number of galaxies involved in the averaging for nodes i/j and the respective standard deviation and n_{ij} is the number of galaxies used in the averaging process for both nodes.

Assuming that shear/ellipticity measurement errors are well-approximated by a Gaussian, we can define the negative log-likelihood for the weak lensing constraints as

$$\chi^2_{\mathrm{WL}}\left(\langle \boldsymbol{\epsilon}_{1/2} \rangle | \boldsymbol{\psi}\right) = \left(\langle \boldsymbol{\epsilon}_{1/2} \rangle - \frac{1}{1-\boldsymbol{\kappa}_{\mathrm{p}}} \mathcal{G}_{1/2} \boldsymbol{\psi}\right)^{\mathrm{T}} \mathcal{C}^{-1}_{\mathrm{WL}} \left(\langle \boldsymbol{\epsilon}_{1/2} \rangle - \frac{1}{1-\boldsymbol{\kappa}_{\mathrm{p}}} \mathcal{G}_{1/2} \boldsymbol{\psi}\right), \tag{3.26}$$

where $G_{1/2}$ is the appropriate finite differencing operator for the respective component of the complex shear and $\langle \epsilon_{1/2} \rangle$ are the corresponding components of the measured average ellipticity. Note that, in order to remain quadratic in ψ, the convergence κ_{p}, necessary to convert shear to reduced shear, is based on a prior and subsequently updated iteratively (see section 3.6).

3.3.2 Strong lensing

In the regime of strong gravitational lensing, the non-linearity of the lens mapping leads to multiple images of a background galaxy appropriately aligned with the cluster. If that galaxy happens to be close to a caustic (see section 1.3), these images blend together and are highly magnified and distorted, forming a giant arc. Recalling that the deflection angle is given by the angular gradient of the lensing potential, the lens equation reads

$$\boldsymbol{\beta}(\boldsymbol{\theta}|\psi) = \boldsymbol{\theta} - \nabla\psi(\boldsymbol{\theta}). \tag{3.27}$$

If now all N_s images stemming from the same source galaxy can be reliably identified, then it is straightforward to define a negative log-likelihood that minimizes the deviation of source plane positions for each of these images

$$\begin{aligned}\chi^2_{\mathrm{MI}}\left(\boldsymbol{\theta}_i|\psi\right) &= \frac{1}{\sigma_\mathrm{s}^2}\left(\boldsymbol{\beta}_i(\boldsymbol{\theta}_i|\psi) - \langle\boldsymbol{\beta}\rangle\left(\psi\right)\right)^2 \\ &= \frac{1}{\sigma_\mathrm{s}^2}\left(\boldsymbol{\theta}_i - \nabla\psi(\boldsymbol{\theta}_i) - \langle\boldsymbol{\beta}\rangle\left(\psi\right)\right)^2,\end{aligned} \tag{3.28}$$

where

$$\langle\boldsymbol{\beta}\rangle\left(\psi\right) = \frac{1}{N_\mathrm{s}}\sum_i^{N_\mathrm{s}}\boldsymbol{\beta}(\boldsymbol{\theta}_i|\psi) \tag{3.29}$$

is the mean source plane position for all images stemming from the same source, given the lensing potential ψ. Then σ_s quantifies the positional error tolerance on the source plane. In terms of FD operators and summed over all members of one multiple image system, this becomes

$$\chi^2_{\mathrm{MI}}\left(\theta|\psi\right) = \sum_i^{N_\mathrm{s}}\frac{1}{\sigma_\mathrm{s}^2}\left[\theta_i - D_{ik}\psi_k - \frac{1}{N_\mathrm{s}}\sum_j^{N_\mathrm{s}}(\theta_j - D_{jk}\psi_k)\right]^2, \tag{3.30}$$

where the index describing the component of coordinate and deflection angle vector is suppressed, as Eq. (3.30) takes the same form for both components.

This approach is usually referred to as source plane minimization and offers the advantage of again being at most quadratic in ψ. The disadvantage however is that it is not a minimization in data space, i.e. the image plane, and uses highly reduced data to begin with, namely the image plane positions of brightness peaks of multiple images as opposed to an actual, or even filtered, brightness map over the field of view. Therefore a likelihood based on source plane minimization does not penalize false positives, which means realisations of ψ (especially in a free-form framework) that minimize this log-likelihood may well produce false additional images without being ruled out. A posteriori checks can of course account for this, and source plane minimization remains in use, due to its appealingly low computational cost.

A more involved approach of using multiple images is often referred to as **source inversion** (Warren and Dye, 2003; Koopmans, 2005). In this scenario,

the extended surface brightness distribution of the images is used to constrain the lens and simultaneously reconstruct the original surface brightness of the source. The clear advantage with respect to the sole usage of image peak positions is the vastly increased amount of data points. But this of course comes at the prize of computational complexity. As we don't expect all points in all images to map back to one singe point on the source plane, reconstructing the source properly is actually non-optional in this approach, which extends the dimensionality of the reconstruction problem by the number of parameters used to model the source. If successful, this on the other hand allows to constrain the morphology of the high-redshift source galaxy. In general the log-likelihood then takes a form

$$\chi^2_{\mathrm{SI}}(\boldsymbol{d}|\boldsymbol{\psi},\boldsymbol{s}) = \left(\boldsymbol{m}\left(\boldsymbol{\psi},\boldsymbol{s}\right) - \boldsymbol{d}\right)^{\mathrm{T}} \mathcal{C}_{\mathrm{SI}}^{-1} \left(\boldsymbol{m}(\boldsymbol{\psi},\boldsymbol{s}) - \boldsymbol{d}\right), \tag{3.31}$$

where $\boldsymbol{d}$ represents the (pixelated) surface brightness data of lensed images with the data covariance $\mathcal{C}_{\mathrm{SI}}$. The response $\boldsymbol{m}(\boldsymbol{\psi},\boldsymbol{s})$ is a function of both the lensing potential and the source surface brightness parameters $\boldsymbol{s}$. In the context of galactic lenses, using a parametric lens model with free-form corrections and a free-form source model, source inversion is for example successfully employed to detect substructures or foreground structures (Vegetti and Koopmans, 2009; Vegetti et al., 2012). In a line-search-like[6] iteration scheme, first a parametric lens model would be reconstructed, and then pixelated correction to the lens potential and the free-form source model are constrained in a maximum-posterior regression with regularization priors for both. Their approach shares many design similarities with the free-form reconstruction presented here and I will now describe how it could be adapted to augment the present framework with source inversion capabilities.

The response is expressed as a linear operation applied to to a block vector comprised of both the linearised free-form corrections to the lensing potential and the source surface brightness $\boldsymbol{r}^{\mathrm{T}} = (\boldsymbol{s},\delta\boldsymbol{\psi})^{\mathrm{T}}$ such that the negative log-likelihood reads

$$\chi^2_{\mathrm{SI}}(\boldsymbol{d}|\boldsymbol{r}) = \left(\mathcal{M}\boldsymbol{r} - \boldsymbol{d}\right)^{\mathrm{T}} \mathcal{C}_{\mathrm{SI}}^{-1} \left(\mathcal{M}\boldsymbol{r} - \boldsymbol{d}\right). \tag{3.32}$$

The elements of the response operator

$$\mathcal{M} = \mathcal{B}\left(\mathcal{L} - \mathcal{D}_s\mathcal{D}_\psi\right) \tag{3.33}$$

are the blurring induces by the instrumental point spread function (PSF) $\mathcal{B}$, the lens operator $\mathcal{L}$ and the correction term $\mathcal{D}_s\mathcal{D}_\psi$, which will be explained below. The lensing operator encodes the mapping between the lens/data and source planes and acts on the source part of $\boldsymbol{r}$. The original authors chose a regular Cartesian grid for the lens plane based on the pixels of the observational data and a Delaunay triangulation for the source plane. This choice of an unstructured mesh is motivated by the fact that the lens mapping induces significant

[6] in line-search optimization, parts of the full parameter space of a model are alternately kept fixed while others are optimized

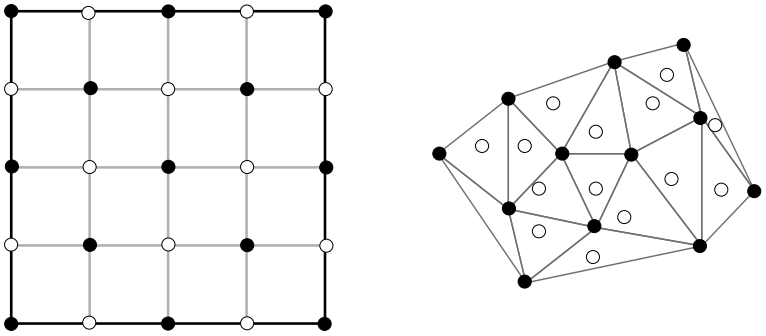

FIGURE 3.5: Illustration of the support structures in (Vegetti and Koopmans, 2009); Half of the nodes of the regular grid representing the image plane on the left get mapped to the source plane (right) and become nodes of the Delaunay triangulation there. The other half, when mapped, falls within the triangular cells.

distortions such that it is a priori unclear where on the source plane high resolution is required. By defining the nodes of a Delaunay mesh as a lensed-back subset of the lens plane pixels, the source plane resolution automatically adapts as reconstruction iterations progress. In such a scenario and if say ever odd-numbered pixel in the data domain is cast back to the source plane to define the nodes of the Delaunay mesh, then every even-numbered pixel would land within one triangular cell of that source mesh[7] and its assigned surface brightness value would be the result of simple linear interpolation based on the three surrounding nodes. This is exactly the mapping encoded in $\mathcal{L}$, illustrated in Figure 3.5. It is a $n_d \times n_s$ matrix, where n_d is the number of data pixels and n_s the number of source nodes. Every row has either one non-zero entry for the corresponding source node, or three entries for the linear interpolation weights if the respective data pixel is not connected to a source node by the lens mapping.

In the present framework both lensing potential and source would be represented by mesh-free domains, a transition that is relatively straightforward. Mesh-free support has the same desired feature of adaptive resolution as the Delaunay triangulation and as section 3.2 showed, interpolation and differentiation are readily implemented with a pre-defined number of nearest neighbours then taking the role of the nodes of the Delaunay cell in both. The next and most obvious change is to drop the parametric lens model and promote the corrections on the lensing potential to be the entire lens model. They are subject to the correction term $\mathcal{D}_s\mathcal{D}_\psi$, which is motivated as follows (Koopmans, 2005). After one iteration of the reconstruction (ignoring blurring for now), the discrepancy between data and model is

$$\delta \boldsymbol{d} = \boldsymbol{d} - \mathcal{L}\boldsymbol{s}. \tag{3.34}$$

[7] or may fall outside the the mesh, but most Delaunay implementations account for such cases by defining an encompassing "infinite" cell

If this discrepancy is due to an incorrect lens mapping and not an incorrect source model, the goal is to correct the lensing potential, such that this discrepancy is minimized. Assuming continuous fields for a moment, we can write

$$\delta d(\boldsymbol{\theta}) = s(\boldsymbol{\beta} + \delta\boldsymbol{\beta}) - s(\boldsymbol{\beta}) \approx \nabla_\beta s(\boldsymbol{\beta})\delta\boldsymbol{\beta}, \tag{3.35}$$

i.e. the discrepancy is caused by a shift in the source plane coordinate $\delta\boldsymbol{\beta}$, such that the wrong surface brightness value gets assigned to the image plane position $\boldsymbol{\theta}$. This, to first order, can be re-expressed using the gradient of the surface brightness. Using the lens equation (1.55), the shift in source coordinate can be associated to a correction in the deflection angle, and thus the gradient of the lensing potential and we arrive at

$$\delta d(\boldsymbol{\theta}) \approx -\nabla_\beta s(\boldsymbol{\beta})\nabla_\theta \delta\psi(\boldsymbol{\theta}). \tag{3.36}$$

In the discrete and linearised response above, the gradient of the source plane surface brightness is expressed by the matrix $\mathcal{D}_s$, whereas the gradient operator on the lens plane is expressed by $\mathcal{D}_\psi$, such that their product, applied to the potential part of $\boldsymbol{r}$ returns the equivalent of Eq. (3.36). As with the non-linearities in the weak lensing contribution, the lensing operator $\mathcal{L}$ as well as the surface brightness gradient $\mathcal{D}_s$ have to be iteratively updated, starting from a reasonable prior. Experiments to incorporate such a source inversion technique in our framework are ongoing, but complicated by the fact, that a purely free-form lens reconstruction can much easier run into unphysical false minima of the log-posterior, caused by numerical artefacts or noise, than a method that only allows linear corrections to a parametric model that gets repeatedly recomputed throughout the reconstruction.

The third method to use strong lensing as a constraint in the present framework is by approximating the positions of critical curves on the image plane by the positions of giant tangential arcs. The extreme magnifications necessary for multiple images to blend together occurs only for sources close to caustics on the source plane. Their images are therefore expected to lie close to the corresponding critical curves, even though they usually don't trace them exactly but rather cross them. Nevertheless, they are in general a reasonable indicator for the position of critical curves. On these critical curves, the determinant of the lensing Jacobian formally has to vanish as the magnification μ formally becomes infinite, which motivates the negative log-likelihood

$$\chi^2_{\mathrm{CC}}(\boldsymbol{\mu}_i|\boldsymbol{\psi}) = \frac{(\det \mathcal{A})^2}{\sigma_i^2} = \frac{\left[(1-\kappa)^2 - \gamma_1^2 - \gamma_2^2\right]}{\sigma_i^2}, \tag{3.37}$$

where $\boldsymbol{\theta}_i$ is the position of a point of the image plane that is part of a giant tangential arc and σ_i is the positional uncertainty. As before, the log-likelihood is kept quadratic in the parameters by iteratively updating the non-quadratic contribution. Expressed in terms of linear operators applied to the lensing

potential and summed over all points belonging to an arc, the log-likelihood thus reads

$$\chi^2_{\mathrm{CC}}(\mu|\boldsymbol{\psi}) = \sum_i \frac{\det \mathcal{A}_{i,\mathrm{p}}}{\sigma_i} \left[\left(1 - K_{ij}\psi_j\right)^2 - \left(G_{1,ij}\psi_j\right)^2 - \left(G_{2,ij}\psi_j\right)^2 \right], \tag{3.38}$$

with $\det \mathcal{A}_{i,\mathrm{p}}$ based on a previous iteration. The degree to which a giant arc actually traces the critical curve depends on the shape and (sub-) structure of the lens and of course on the resolution of the observation, but they are generally expected to be sufficiently close on the image plane.

3.4 Regularisation

As already mentioned above, free-form reconstruction methods are particularly prone to the problem of overfitting. To avoid this inadvertent modelling of noise patterns as part of the lensing potential, a prior/regularisation term that penalizes deviations from a smooth potential is required. This can be achieved by minimizing the difference between the local curvature of the potential, i.e. its Hessian, and that of a smooth moving prior. Since the Hessian of a function is the transpose Jacobian of its gradient, the Hessian of the lensing potential by construction is just $\mathbb{1} - \mathcal{A}$, and thus given by the convergence and the shear. We opt to minimize the difference between the reconstructed and prior convergence and shear in a Gaussian approach and end up with the terms

$$\chi^2_{\mathrm{reg},\kappa}(\boldsymbol{\psi}|\boldsymbol{\alpha}_\kappa) = \left(\boldsymbol{\kappa}_\mathrm{p} - \mathcal{K}\boldsymbol{\psi}\right)^\mathrm{T} \mathcal{H}_\kappa \left(\boldsymbol{\kappa}_\mathrm{p} - \mathcal{K}\boldsymbol{\psi}\right) \tag{3.39}$$

$$\chi^2_{\mathrm{reg},\gamma}(\boldsymbol{\psi}|\boldsymbol{\alpha}_\gamma) = \left(\boldsymbol{\gamma}_\mathrm{p} - \mathcal{G}\boldsymbol{\psi}\right)^\mathrm{T} \mathcal{H}_\gamma \left(\boldsymbol{\gamma}_\mathrm{p} - \mathcal{G}\boldsymbol{\psi}\right), \tag{3.40}$$

where $\mathcal{H}_\kappa$ and $\mathcal{H}_\gamma$ are diagonal matrices with entries $\boldsymbol{\alpha}_\kappa$ and $\boldsymbol{\alpha}_\gamma$ on their respective main diagonals. They set the regularisation strength, i.e. the relative weight of the regularisation term with respect to the data likelihoods, potentially node by node. There is no distinct way to determine the regularisation strength from first principles. It's mostly set using guidance from experience or optimised for, using simulated data. The moving priors κ_p and γ_p are taken from a previous iteration (again see section 3.6), starting with an initial guess, which here will always be a flat zero, i.e. no lens at all. While this starting point is indeed smooth, the regularisation described here on its own is no guarantee that noise in the data will not drive the reconstructed solution far away from smoothness. This problem will be mitigated by the two-level iteration scheme described in section 3.6. Another possible approach for smooth regularisation, via an entropic term involving a smoothed moving prior, will be presented in the next section. For the overall reconstruction framework presented here, we opt for the least-squares terms (3.39, 3.40) for reasons of computational simplicity on the mesh-free domain.

If the method is augmented by the source inversion techniques discussed above, the parameter space of the reconstruction is enlarged by the source

model and an additional regularisation term for the surface brightness of the source becomes necessary. Various approaches can be found in the literature (Suyu et al., 2006; Tagore and Keeton, 2014) and often again constrain the Hessian of the model in order to avoid overly peaked solutions.

3.5 Adding non-lensing constraints

As the projected gravitational potential ψ is the quantity defining all lensing observables and therefore the quantity this reconstruction framework is built on, any additional non-lensing constraint that is to be incorporated in this framework has to be relatable to ψ. Luckily gravity is a driving force for most astrophysical processes, so a physically motivated relation (rather than an empirical scaling relation) to the Newtonian potential Φ of a cluster can be constructed for a variety of observables. To do so however, it is necessary to overcome a fundamental problem in Astronomy: the physical processes happen in three-dimensional space whereas observations are taken in one projection, on the two-dimensional sky. So in order to obtain an estimate on ψ based on a non-lensing observable via physically motivated relations one has to take a detour. First *deproject* the observed quantity to get an estimate of the underlying three-dimensional distribution, then use *physical connections* to get an estimate on the three-dimensional Newtonian potential and subsequently *project* the latter to arrive at ψ. This section will describe these steps, first introducing the deprojection algorithm used, then detailing the physics involved in connecting X-ray, thermal SZ and kinematic observations to the Newtonian potential. Lastly problems arising from re-projecting discrete and bounded data as well as possible solutions are presented.

At this point it is important to note the following limitation of the presented method. Recalling the introduction of gravitational lensing in section 1.3.1, especially equations (1.49) and (1.51), it becomes obvious that in a general metric theory of gravity, lensing probes the the difference[8] $\Phi - \Psi$ of the two Bardeen potentials, the so-called Weyl potential. The physical processes involving non-relativistic matter described below however only depend on the Newtonian potential Φ, i.e. the potential in the time-time term of the metric element. Only if General Relativity as the specific theory of gravity is assumed, the two potentials are identical up to a sign and thus relativistic and non-relativistic particles probe the same potential. Only in this case observations based on relativistic and non-relativistic particles can be combined in the way it is done here.

Several ways to include the potential estimate based on de- and reprojected data are conceivable, depending on the specifics of the deprojection algorithm and the data at hand. In general the respective log-likelihood term that enters the reconstruction will be of the form

$$\chi^2_{\mathrm{X}}(\bar{\psi}_{\mathrm{X}}|\psi) = (\bar{\psi}_{\mathrm{X}} - \mathcal{P}\psi)^{\mathrm{T}}\, \mathcal{C}_{\mathrm{X}}^{-1}\, (\bar{\psi}_{\mathrm{X}} - \mathcal{P}\psi)\,. \tag{3.41}$$

[8] or sum, depending on sign conventions

$\mathcal{P}$ here denotes any operator that ensures compatibility of the coordinate basis of the model ψ and that of the data-based estimate $\bar{\psi}_X$ and $\mathcal{C}_X$ is the covariance matrix of $\bar{\psi}_X$. The operator $\mathcal{P}$ depends mostly on the specifics of the deprojection method and will be revisited in section 3.6.

3.5.1 Richardson-Lucy deprojection with correct amplitudes

Here I briefly recap the main ideas of the Richardson-Lucy (RL hereafter) deprojection method. For more extensive accounts see Lucy (1974), Lucy (1994), Konrad et al. (2013), and Majer et al. (2016) among others. The RL algorithm is chosen here as its design resonates well with that of the lensing reconstruction. It is a local method to reconstruct the three-dimensional distribution of a physical quantity based on a two-dimensional projection, without assuming a specific physical model. It is however not entirely free-form, since any deprojection method needs assumptions concerning the distribution along the line of sight. In the present case, this is done by assuming that the three-dimension distribution adheres to certain degree of symmetry. For brevity and simplicity we here assume spherical symmetry and thus radial profiles, but the method can be extended to other symmetries as well (Reblinsky, 2000; Majer et al., 2016).
The RL algorithm was developed as a solver for any inverse problem that can be phrased as a Fredholm or Volterra[9] integral equation of the first kind, i.e.

$$g(x) = \int \mathrm{d}\xi K(x|\xi) f(\xi), \tag{3.42}$$

where g is a function accessible to observations while f is the not directly observable target function and $K(x|\xi)$ is the kernel connecting the two. The applications the original authors had in mind were all reducible to cases where g, f and $K(x|\xi)$ are probability density functions (pdf) and therefore adhere to normalization and non-negativity constraints. In that case, Bayes' theorem in Eq. (3.1) states that

$$K(\xi|x) = \frac{K(x|\xi) f(\xi)}{g(x)}, \tag{3.43}$$

where $K(\xi|x)$ is a pdf describing the unknown inverse kernel, such that

$$f(\xi) = \int \mathrm{d}x K(\xi|x) g(x). \tag{3.44}$$

Equations (3.43) and (3.44) suggest an iterative scheme to estimate $f(\xi)$ given $g(x)$ and an initial guess $\tilde{f}_0(\xi)$, by updating $\tilde{f}(\xi)$ in every step using an estimated inverse kernel based on the reprojected result of the previous iteration. This scheme can be expressed as a maximum likelihood approach (Lucy, 1974) and refined by adding an entropic regularization term (Lucy, 1994) and thus qualifies as a maximum entropy method. The resulting algorithm will now be detailed for the case of deprojecting a spherically symmetric field.

[9] Fredholm equations have fixed integration bounds, in Volterra equations they are variable

Assuming a three-dimensional field $f(r)$, a two-dimensional projection of said field can be expressed as

$$g(s) = \int \mathrm{d}r f(r) K(s|r), \tag{3.45}$$

where r and s denote three- and two-dimensional radius respectively[10] and $K(s|r)$ is the projection kernel, which for spherical symmetry reads

$$K(s|r) = \frac{r}{\sqrt{r^2 - s^2}} \Theta(r^2 - s^2). \tag{3.46}$$

Here $\Theta(r^2 - s^2)$ is the Heaviside step function. Given the projected field $g(s)$, the RL-deprojection algorithm infers locally, i.e. for each line of sight, a free-form (i.e. no underlying model is assumed, only symmetry) estimate $\tilde{f}(r)$ of the three dimensional field, by iteratively maximizing the objective functional

$$Q[\tilde{f}] = H[\tilde{f}] + S[\tilde{f}], \tag{3.47}$$

where

$$H[\tilde{f}] = \int \mathrm{d}s \, g(s) \ln \tilde{g}(s) \tag{3.48}$$

is the log-likelihood for a reprojected estimate $\tilde{g}(s)$ and

$$S[\tilde{f}] = -\alpha \int \mathrm{d}r \, \tilde{f}(r) \ln \frac{\tilde{f}(r)}{\pi(r)} \tag{3.49}$$

an entropic regularization with regularization strength α and a smooth moving prior $\pi(r) = \int \mathrm{d}r' P(r|r') \tilde{f}(r)$, where $P(r|r')$ is a appropriately chosen window function.

The explicit form of H stems from the assumption that the measured $g(s)$ are drawn from a Poisson distribution[11] with the mean given by the reprojected estimate $\tilde{g}(s)$, independently for each s. The regularization term S on the other hand penalizes a large Kullback-Leibler divergence between the inferred $\tilde{f}(r)$ and a smoothed version of itself, ensuring a sufficiently smooth final estimate.

Starting from an initial guess for the three-dimensional field $\tilde{f}_0$, each algorithmic step takes the form

$$\begin{aligned}
\tilde{f}_{i+1} &= \tilde{f}_i + \Delta\tilde{f}_i \\
\Delta\tilde{f}_i &= \frac{\delta Q}{\delta \tilde{f}_i} - \int \mathrm{d}r \tilde{f}_i(r) \frac{\delta Q}{\delta \tilde{f}_i} = \Delta H + \Delta S \\
\Delta H &= \tilde{f}_i \left(\int \mathrm{d}s \frac{g(s)}{\tilde{g}_i(s)} K(s|r) - 1 \right) \\
\Delta S &= -\alpha \tilde{f}_i \left(S + \ln \frac{\tilde{f}_i}{\pi_i} + 1 - \int \mathrm{d}r' \frac{\tilde{f}_i(r')}{\pi_i(r')} P(r|r') \right).
\end{aligned} \tag{3.50}$$

[10] $r = \sqrt{s^2 + z^2}$, where z here denotes the coordinate along the line of sight

[11] even though $g(s) \in \mathbb{R}_{\geq 0}$

The algorithm converges usually within only a few iterations, even when the initial guess is significantly off. It preserves normalizations and ensures, within a certain range of regularization strengths α, also non-negativity of all fields and kernels involved. As explained above, for reasons of internal consistency, fields and kernels need to be normalized with respect to their domains, but the correct amplitude and units for the resulting $\tilde{f}(r)$ can be restored as follows.
Assuming that the two dimensional input field and the projection kernel are normalized via

$$g_{\mathrm{n}}(s) = \frac{1}{I_g} g(s), \tag{3.51}$$

$$K_{\mathrm{n}}(s|r) = \frac{1}{I_K(r)} K(s|r) \tag{3.52}$$

$$I_g = \int \mathrm{d}s g(s) \tag{3.53}$$

$$I_K(r) = \int \mathrm{d}s K(s|r), \tag{3.54}$$

then inserting Eq. (3.52) and (3.45) in Eq. (3.51) and integration over the two-dimensional domain yields

$$\underbrace{\int \mathrm{d}s g_{\mathrm{n}}(s)}_{=1} = \int \mathrm{d}r \frac{I_K(r)}{I_g} f(r) \underbrace{\int \mathrm{d}s K_{\mathrm{n}}(s|r)}_{=1}, \tag{3.55}$$

such that

$$\int \mathrm{d}r \tilde{f}_{\mathrm{n}}(r) = 1, \quad \tilde{f}_{\mathrm{n}}(r) = \frac{I_K(r)}{I_g} f(r). \tag{3.56}$$

So we can recover the units and scale of the three-dimensional field via

$$\tilde{f}(r) = \frac{I_g}{I_K(r)} \tilde{f}_{\mathrm{n}}(r) \tag{3.57}$$

from the normalized output of the RL-algorithm $\tilde{f}_{\mathrm{n}}(r)$. Thus equipped with a deprojection method, we move on to connecting the deprojected quantity to the Newtonian potential.

3.5.2 X-ray

In the case of X-ray observations, the observed two-dimensional data are counts of X-ray photons hitting a detector. The number of photons detected N_{X} is connected to the surface brightness via

$$N_{\mathrm{X}}(s) = \frac{\Delta t A_{\mathrm{eff}}}{\hbar \bar{\omega}} S_{\mathrm{X}}(s), \tag{3.58}$$

where Δt is the exposure time, A_{eff} the effective area of the detector and $\hbar\bar{\omega}$ the average energy of the photons. Recalling section 2.1.3, the X-ray surface brightness can be expressed as the line-of-sight projection of the frequency-integrated X-ray emissivity of the ICM

$$j_{\text{X}}(r) = \text{C}\rho^2(r)\sqrt{\text{k}_\text{B}T(r)}, \tag{3.59}$$

where $\rho(r)$ and $T(r)$ are gas density and temperature respectively and where we have collected the constant terms in Eq. (2.7) in the overall constant C, which is approximately

$$C \approx 6.89 \cdot 10^{23} \, \frac{\text{erg}\,\text{cm}^3}{\text{g}^2\,\text{s}\sqrt{\text{eV}}}$$

for a fully ionized hydrogen plasma[12]. As we assume spherical symmetry, the density and temperature and consequently the emissivity are treated as functions of three-dimensional radius r only.

So, by translating the photon counts to the surface brightness and deprojecting the latter, we arrive at the frequency-integrated emissivity. To get an estimate for the gravitational potential of the cluster from this, some assumptions about the state of the ICM have to be made. Here we opt for

- an ideal gas law
$$k_\text{B}T = \bar{m}\frac{P}{\rho}, \tag{3.60}$$
allowing to express the ICM temperature as a function of density and pressure P.

- a polytropic stratification
$$\frac{P}{P_0} = \left(\frac{\rho}{\rho_0}\right)^\gamma, \tag{3.61}$$
with polytropic index γ, connecting pressure and density

- and (approximate) hydrostatic equilibrium
$$\nabla P = -\rho\nabla\Phi \tag{3.62}$$
connecting pressure and density to the Newtonian potential

(Konrad et al., 2013; Tchernin et al., 2018). As mentioned in chapter 2, at least the last assumption is not necessarily valid for every cluster but nevertheless often used. Studies suggest that the deviations from hydrostatic equilibrium are most relevant in the innermost cores and the outskirts of clusters as well as in recent mergers (Shi and Komatsu, 2014; Biffi et al., 2016). Also polytropic

[12] $Z = 1$, $\bar{m} = \frac{1}{2}m_{\text{proton}}$

stratification can not necessarily be safely assumed for the centres of clusters, as they may host cooling cores (Markevitch, 1998). The locality of the RL deprojection method however allows us to restrict ourselves to intermediate ranges in radius, which should minimize the effects of hydrostatic bias and localised cooling flows, by neglecting the areas of the cluster, where these biases dominate.

Introducing the dimensionless gravitational potential

$$\varphi(r) = -\frac{\gamma - 1}{c_{s,0}^2} \left(\Phi(r) - \Phi_{\mathrm{cut}}\right) \tag{3.63}$$

with the sound speed at some reference radius r_0

$$c_{s,0}^2 = \gamma \frac{P_0}{\rho_0} = \gamma \frac{k_B T_0}{\bar{m}} \tag{3.64}$$

and the potential at a cut-off radius $\Phi_{\mathrm{cut}} \equiv \Phi(r_{\mathrm{cut}})$, the emissivity can be expressed as a function of that dimensionless gravitational potential

$$j_X(r) = C \rho_0^2 T_0^{1/2} \varphi^{(3+\gamma)/(2\gamma-2)}(r). \tag{3.65}$$

T_0 and ρ_0 are temperature and density at the reference radius r_0 respectively.

Thus, using the aforementioned assumptions, an estimate of the polytropic index and a measurement of the X-ray temperature of the ICM are needed to infer the three dimensional gravitational potential $\Phi(r)$ from the observed X-ray surface brightness.

3.5.3 Thermal SZ

As introduced in section 2.1.3, the two-dimensional observable of the thermal SZ effect is the intensity change of CMB photons with respect to the Planck spectrum

$$\begin{aligned} \frac{\Delta I_{\mathrm{tSZ}}(x,s)}{B_\omega} = g(x) y(s) &= g(x) \frac{k_B \sigma_T}{m_e c^2} \int \mathrm{d}\chi T_e(s,\chi) n_e(s,\chi) \\ &= \int \mathrm{d}\chi P_{\mathrm{eff}}(s,\chi), \end{aligned} \tag{3.66}$$

where we have defined the effective pressure P_{eff} (Majer et al., 2016). Thus, deprojecting the spectral distortion, we can determine the effective pressure. Using the same assumptions as in the X-ray case above, that is

- ideal gas law
- polytropic stratification
- approximate hydrostatic equilibrium,

we can again connect density, temperature and hence the effective pressure to the dimensionless gravitational potential and arrive at the relation

$$P_{\rm eff} = g(x)\frac{k_{\rm B}\sigma_{\rm T}}{m_e c^2}\rho_0 T_0 \varphi^{\gamma/(\gamma-1)}. \tag{3.67}$$

And again, we additionally need an estimate of the polytropic index and the ICM temperature to recover the Newtonian potential.

3.5.4 Kinematics

As shown in section 2.1.3, the radial velocity dispersion of galaxy members within the cluster can be connected to the Newtonian potential via the Jeans equation (2.5). The actual observable is the density-weighted projection of that dispersion. We can identify the density-weighted radial velocity dispersion with an effective pressure $P_{\rm eff} = n\sigma_{rr}^2$, which illustrates the formal similarity of the Jeans and the hydrostatic equation, as the former now reads

$$\frac{\partial P_{\rm eff}}{\partial r} + 2\beta_{\rm ani}\frac{P_{\rm eff}}{r} = -n\frac{\partial \Phi}{\partial r}, \tag{3.68}$$

where we introduced the common anisotropy parameter $\beta_{\rm ani} = 1 - \frac{\sigma_{\theta\theta}^2}{\sigma_{rr}^2}$ and set $\sigma_{\theta\theta}^2 = \sigma_{\phi\phi}^2$ due to the isotropy assumption (Binney and Tremaine, 1987). Numerical tests in Sarli et al. (2014) indicate that a polytropic stratification for this effective galaxy pressure can be assumed, such that we can again close the system of equations and connect the effective pressure uniquely to the dimensionless gravitational potential via

$$\begin{aligned} p_{\rm eff}^{\eta}(r) = &-\varphi(r) + \exp\left(-2\int_{r_0}^{r} {\rm d}x \frac{\eta\beta_{\rm ani}}{x}\right) \\ &+ 2\int_{r_0}^{r} {\rm d}y \frac{\eta\beta_{\rm ani}}{y}\varphi(y)\exp\left(2\int_{r}^{y} {\rm d}x \frac{\eta\beta_{\rm ani}}{x}\right), \end{aligned} \tag{3.69}$$

with

$$\eta = \frac{\gamma-1}{\gamma}, \quad p_{\rm eff} = \frac{P_{\rm eff}}{P_0}, \quad \varphi = \frac{n_0\eta}{P_0}\Phi,$$

and the polytropic index γ. This connection is more complicated than in the cases above, due to the anisotropy term in the Jeans equation. In fact Eq. (3.69) is a Volterra integral of the 2nd kind and needs to be solved numerically. Despite this, the general idea remains the same and estimates of the polytropic index and the effective pressure at some reference radius are needed to obtain the Newtonian potential.

3.5.5 Artefact treatment

Now that we are able to infer the three-dimensional Newtonian potential from non-lensing observations we just need to project it in order to arrive at an

estimate of the lensing potential that can be incorporated into our reconstruction framework. There is however a complication. The discrete and truncated nature of our data-based potential estimate can introduce numerical artefacts during the projection that need to be taken care of. I will first introduce the origin of such artefacts and then present means to amend them.

Origin of artefacts

In our present scenario of a thin lens adhering to spherical symmetry, the projection integral for the lensing potential formally is given by

$$\psi(s) = \frac{2}{c^2}\frac{D_{\mathrm{ls}}}{D_{\mathrm{l}}D_{\mathrm{s}}}\int_s^\infty \mathrm{d}r \frac{r}{\sqrt{r^2-s^2}}\Phi(r), \tag{3.70}$$

where, in comparison to Eq. (1.66), we replaced the line-of-sight integral with an integral over the three-dimensional radius using the projection kernel defined in Eq. (3.46). But since the gravitational potential is based on finite data, for example the X-ray surface brightness within a finite field of view, it is only known up to a certain radius r_{data}. The projection integral thus is truncated at r_{data}

$$\psi\,(s) \approx \frac{2}{c^2}\frac{D_{\mathrm{ls}}}{D_{\mathrm{l}}D_{\mathrm{s}}}\int_s^{r_{\mathrm{data}}} \mathrm{d}r \frac{r}{\sqrt{r^2-s^2}}\Phi(r). \tag{3.71}$$

The applicability of this approximation[13] critically depends on the choice of r_{data}. For large enough fields of view, the gravitational potential may drop off to small enough absolute values within r_{data} such that the region outside does not significantly contribute to the integral and therefore can safely be neglected. If the field of view is however comparable to the virial radius of the cluster, these neglected contributions become increasingly important and simply cutting them away introduces artefacts that severly alter the radial curvature of the resulting lensing potential and thus systematically bias conclusions on the mass distribution in the lens. Figure 3.6 illustrates this effect. It shows the difference of the projected gravitational potential of a NFW halo according to Eq. (3.71) from the analytically known lensing potential. If the integral is truncated at the virial radius, the result significantly deviates from the truth on all radii. With increasing upper bound, the differences become smaller and restricted to larger and larger radii. This effect is less pronounced for functions that fall off steeper as can be seen in Figure 3.7, where the same procedure is shown for the projected density versus the analytically known convergence. Since the X-ray emissivity of the ICM is quadratic in the gas density and therefore falls off even steeper, it is not surprising, that truncation

[13] It is perhaps noteworthy that the truncation is not the only approximation. Of course data, and thus the estimated gravitational potential, is discrete which means the integral in the projection is replaced by a sum. This however proves to have little influence on the result as even a simple midpoint Riemann sum produces good results once the truncation radius is large enough.

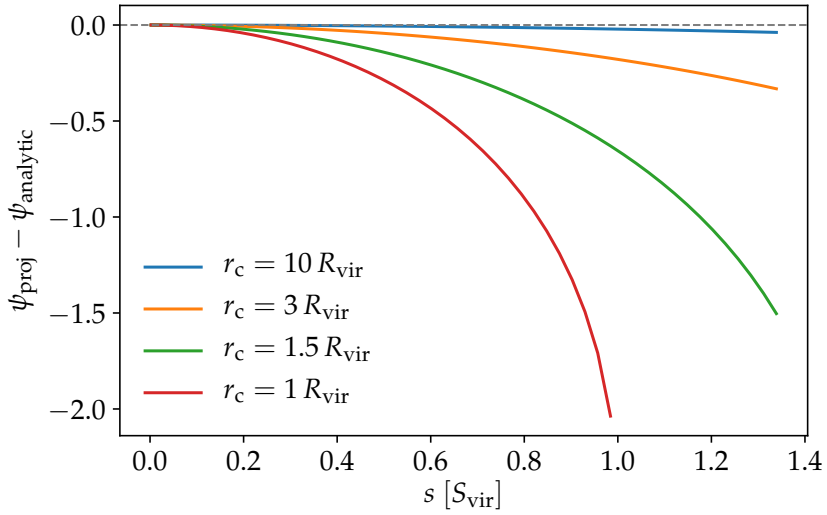

FIGURE 3.6: Difference between projections of the gravitational potential of a NFW halo, truncated at different multiples of the virial radius, and the analytic lensing potential of a NFW lens, as a function of two-simensional radius s; all curves shifted to coincide at the innermost radius

artefacts usually do not interfere with the RL deprojection algorithm even though it includes a reprojection in every iteration.

Analytic continuation

To amend these artefacts in cases where radial extent is not large enough to ignore them, we chose to introduce an analytic continuation of the 3D gravitational potential outside of r_{data}. Such an extrapolation needs to be justified and treated with care. Several functional forms are possible to reasonably describe the gravitational potential outside the cluster and most of them formally diverge when integrated up to infinite radius, so still an upper bound for this continuation has to be defined. Both the choice of functional form and upper bound of the continuation introduce some level of arbitrariness to the method, but are necessary to counteract the biasing effect of projection artefacts. This leads to the adjusted projection integral

$$\psi(s) \approx \frac{2}{c^2}\frac{D_{\mathrm{ls}}}{D_{\mathrm{l}}D_{\mathrm{s}}}\left[\int_s^{r_{\mathrm{data}}} \mathrm{d}r \frac{r}{\sqrt{r^2-s^2}}\Phi(r) + \int_{r_{\mathrm{data}}}^{r_{\mathrm{cut}}} \mathrm{d}r \frac{r}{\sqrt{r^2-s^2}} f_{\mathrm{cont}}(r)\right], \quad (3.72)$$

where $f_{\mathrm{cont}}(r)$ is the analytic continuation function and r_{cut} is the upper bound of the integration outside r_{data}.

The simplest, and in a way most conservative choice is to assume that there is no significant mass contribution outside of r_{data} and that therefore the

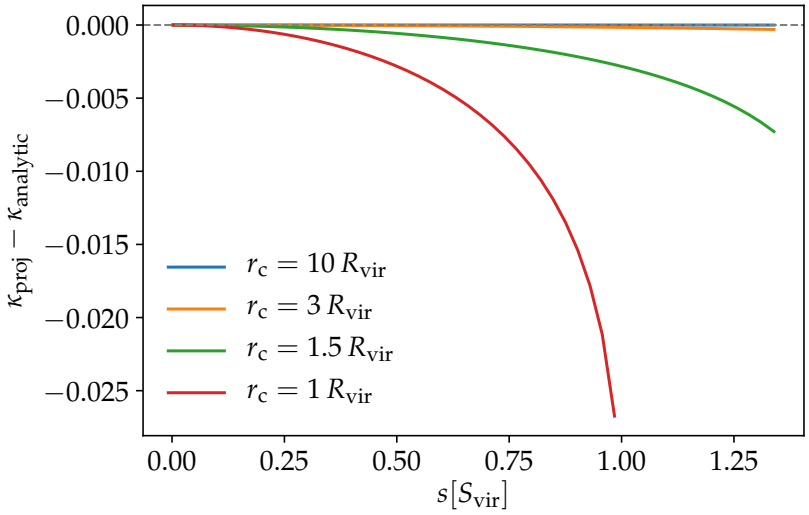

FIGURE 3.7: Difference between projections of the density of a NFW halo, truncated at different multiples of the virial radius, and the analytic lensing convergence of a NFW lens, as a function of two-simensional radius s; all curves shifted to coincide at the innermost radius

potential in this radial regime is dominated by the mass enclosed within r_{data}. In this case the continuation takes the form

$$f_{\mathrm{cont}}(r) = \frac{A}{r} + B, \tag{3.73}$$

where A,B are fixed by smoothness conditions, i.e. we demand that

$$\Phi(r_{\mathrm{data}}) \stackrel{!}{=} f(r_{\mathrm{data}}) \tag{3.74}$$

$$\frac{\partial \Phi}{\partial r}(r_{\mathrm{data}}) \stackrel{!}{=} \frac{\partial f_{\mathrm{cont}}}{\partial r}(r_{\mathrm{data}}). \tag{3.75}$$

While the assumption, that there is no more mass outside r_{data}, most likely does not hold in a strict sense, tests with hydrodynamical simulations show that it can be sufficient to reconcile reconstructed and true projected potential (see section 4.2).

3.6 Implementation

After introducing a classification system of reconstruction methods and positioning our own method within that system, I discussed the various data we can use to constrain the lensing potential of a galaxy cluster. In the following, I finally present the current specific implementation of this reconstruction framework. From here on, non-lensing constraints will be restricted to X-ray data, deprojected assuming spherical symmetry, as that example already highlights

all the features and steps relevant for the combination of non-lensing and lensing data in our free-form reconstruction. First I will detail some preparatory steps that need to be taken in order to render X-ray surface brightness data usable for our purposed an then I will present the implemented reconstruction code itself, SAWLENS2. In the last section, I will go over the pipeline employed to analyse the outcome of the reconstruction and the ways to estimate and assign errors to it.

3.6.1 Preparations

In order to use X-ray data in SAWLENS2 as indicated by Eq. (3.41), we so far still lack an estimate of the noise and correlation properties of the data, and how this affects the estimated lensing potential. In other words, we have to determine the covariance matrix $\mathcal{C}_\chi$. To do so, we produce a number of realisations of the data, according to the expected probability distribution of the data, and then subject these realisations to the de- and reprojection procedure outlined above, to arrive at a set of realisations of the estimated lensing potential $\bar{\psi}_\chi$. Based on these realisations we then can compute the covariance matrix.

As the X-ray surface brightness is measured counting photons hitting a detector and since we can always rebin or coarse-grain the data such that the point spread function and pixel correlations due to detector systematics can be ignored, we can draw realisations in the present case from a Poisson distribution, where the mean is given by the actual observation.

Since a large number of realisations is needed to accurately infer the covariance matrix and since we are only interested in the radial profiles of surface brightness anyway (due to the assumption of spherical symmetry in the deprojection step), it is convenient to first reduce the observed photon count map to a cumulative profile, that still adheres to the same noise properties but requires much less space and computational time. To this end, we define the center of the count map to be the peak of the surface brightness (excluding obvious artefacts and point sources) and draw a number of evenly or logarithmically spaced radial bins around that center, before summing up the counts of all pixels in each of those bins. We then draw the above mentioned Poisson realisations, as this profile of sums is still adhering to the same statistics as the counts. After the realisations are created, we divide the number of counts in each bin by the number of pixels in that bin to arrive at realisations of the azimuthally averaged radial count profile. With the average photon energy, the exposure time and the effective area of the detector, we can compute the surface brightness profiles from the count profiles and are able to deproject them to obtain a set of realisations of the X-ray integrated emissivity. All steps so far are easily implemented in any scripting language and template python code is provided and packaged with the deprojection implementation, which is written in C++ for performance reasons.

To compute the gravitational potential from that, we now require an estimate of the polytropic index γ and of the temperature of the ICM (see section

3.5). The polytropic index can be determined by plotting the frequency integrated X-ray emissivity against the ICM pressure, based on deprojected thermal SZ observations, on the same domain, since both relate to the dimensionless potential with different powers of γ. This plot also immediately identifies the radial regime in which the assumption of polytropic stratification can be safely used, as deviations from it result in a broken power-law behaviour. The application of our method to a realistic hydrodynamical simulation will illustrate this further.

The temperature of the ICM can be inferred from the shape of the X-ray spectrum and for our purposed it is only important that it is measured at the same radius or averaged over the same domain as the normalisation of the emissivity. As an example, if ρ_0, T_0 are supposed to be fixed at a certain radius r_0, we divide the emissivity by its value at r_0 to translate it to the dimensionless potential according to Eq. (3.65). In order to determine the dimensional Newtonian potential we then have to obtain the ICM temperature and thus the sound speed at that same radius r_0. Where that radius is, is largely irrelevant as long as internal consistency is ensured.

For the projection we assume an analytic continuation according to Eqs. (3.72, 3.73) up to a cut-off radius that is a multiple of r_{data}. As mentioned above, the parameters of the analytic continuation are set by smoothness conditions, i.e. by demanding that the values of $f_{\mathrm{cont.}}(r_{\mathrm{data}})$ and its first derivative coincide with the respective values in the deprojected profile. In order to ensure that especially the finite-differencing-based radial derivative for the latter is still driven by data rather than noise at r_{data} we employ a series of simple sanity checks.

We estimate the signal to noise ratio for the enclosed mass as a function of radius

$$G M_{\mathrm{enclosed}}(r) = r^2 \frac{\partial \Phi(r)}{\partial r} \tag{3.76}$$

from Poisson realizations of the X-ray data, as this directly sets the slope of the continuation A. Figure 3.8 shows an example for a mock NFW halo (see also section 4.1).
Additionally we check for unphysical behavior by testing the non-negativity of the Laplacian $\Delta\Phi(r)$, which is of course proportional to the density via Poissons equation.
If the outer bins fail to meet these criteria, the analytic continuation parameters would likely be set by random fluctuations and introduce new artifacts rather than correcting those induced by truncating the projection integral. Therefore, if necessary we cut the profile at the outermost radial bin where both conditions are still satisfied and start the continuation there.

3.6.2 SaWLens2

SaWLens2 is a modular C++ code framework to allow reconstructions of the projected potential of a gravitational lens based on mesh-free support. Its development lineage traces back to Bartelmann et al. (1996) and Cacciato et al.

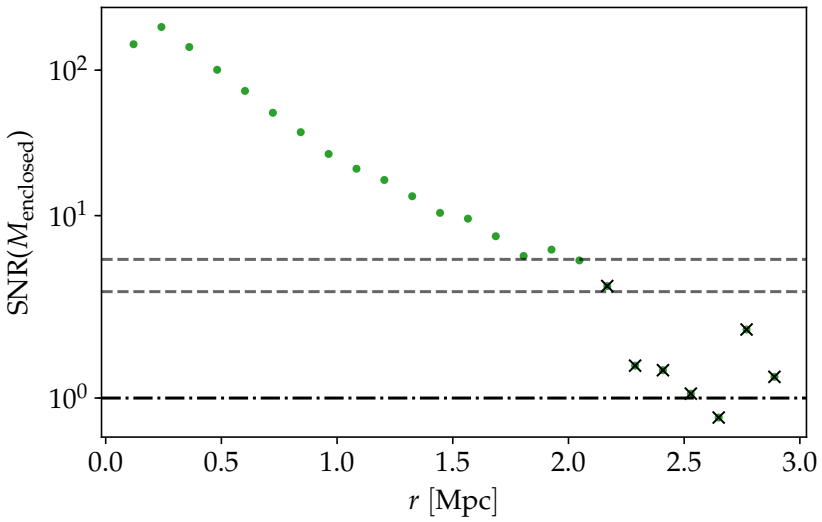

FIGURE 3.8: signal-to-noise ratio for the enclosed mass of a mock NFW halo (see section 4.1) as a function of radial bin. Horizontal lines denote signal-to-noise values of 1, 3 and 5 respectively. Crossed-out bins fail the sanity checks and are excluded from the analysis.

(2006), its first version under this name, then still using adaptive regular grids, was presented and applied to the CLASH sample of massive galaxy clusters in Merten et al. (2009) and Postman et al. (2012) and its current overhauled mesh-free version was first presented in Merten (2016). A very similar approach is described in Bradač et al. (2005). SAWLENS2 is a polymorphic library of classes and routines, falling into three broad groups, the `field node builders`, the `field` and the `system builders`, as well as a number of auxiliary routines to manage inputs, outputs and the like. `Field node builders` translate different data sets into a format usable in SAWLENS2, and create a mesh-free set of support nodes based on this. In principle every data set used in a reconstruction can contribute with its own mesh-free domain. The field class combines these into a unified set of nodes that defines the support for the reconstructed potential. Further, the `field` implements all of the necessary book keeping in the course of the reconstruction. Finally, the `system builders` implement the contributions to the linear system of equations constraining the lensing potential, based on the log-likelihood terms defined above.The LSE is then solved for the lensing potential by the `field` class, using standard `CLAPACK`[14] linear algebra routines. Finally, the resulting lensing potential at formally infinite source redshift (see section 3.2) is converted to the convergence κ at a specified redshift by the auxiliary class `lens`. Figure 3.9 schematically illustrates the interplay of these classes. A ready-to-use driver for cluster reconstructions is packaged with the code, but the modular and polymorphic structure of the code allows for implementations tailored to the use case at hand.

[14] `https://www.netlib.org/clapack/`

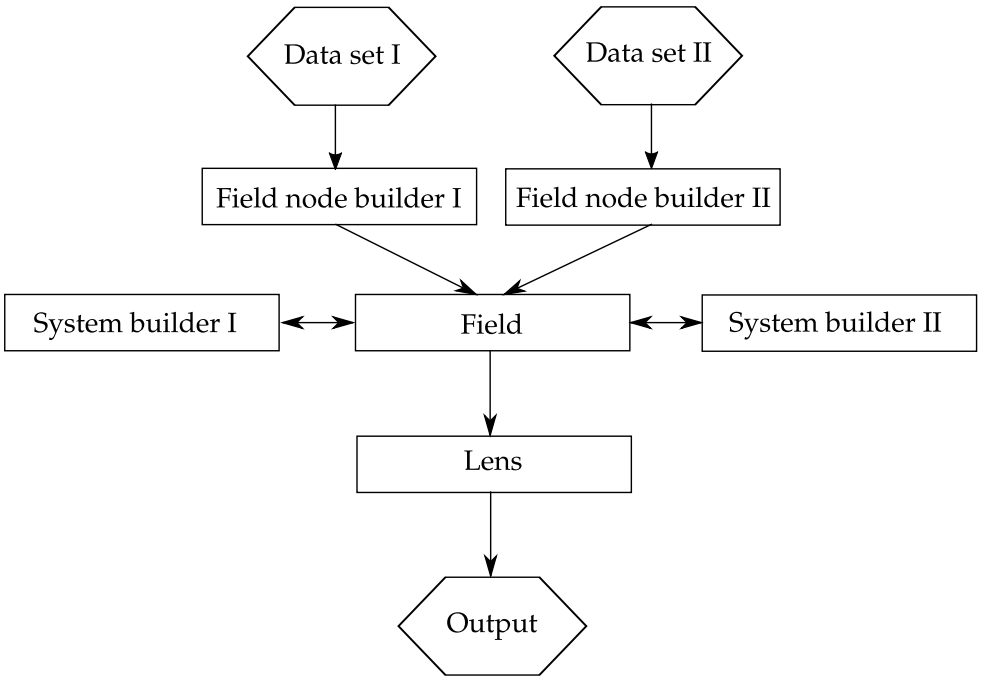

FIGURE 3.9: Schematic illustration of the basic classes in SAWLENS2, as well as their interplay;

In the following I will describe the specific choice and management of the mesh-free support chosen here, before detailing the iteration scheme employed in SAWLENS2. Finally I will introduce the explicit implementation of the X-ray log-likelihood used in the applications in the next chapter. As SAWLENS2 started out as a lensing reconstruction framework, many of its key design features, like the support structure and iteration scheme, are motivated by the peculiarities of especially weak lensing.

Mesh-free support

The mesh-free domain as well as the RBF-based interpolation and differentiation is handled using the separate library `libmfree`[15], which implements all RBFs listed in Table 3.1 and all FD operators necessary for lensing purposes, as listed in Eqs. (3.15-3.18), as well as regular derivatives up to 2nd order in one, two and three dimensions. The shape parameter of the RBFs can be optimized against a selection of analytic test functions on any given mesh-free domain.

We chose the support structure of our reconstruction as follows. For the weak lensing contribution the `field node builder` divides the shear catalogue up in a number of regularly spaced cells on the sky. Within each cell, it randomly selects a galaxy and the coordinates of this galaxy are taken to be the coordinates of a support node. It then identifies the nearest neighbouring galaxies and assigns their mean shear to that node and computes the covariance matrix according to Eq. (3.25). This most basic approach is well-suited for evenly distributed shear catalogues as the ones we will use in the next

[15] `https://bitbucket.org/jmerten82/libmfree`

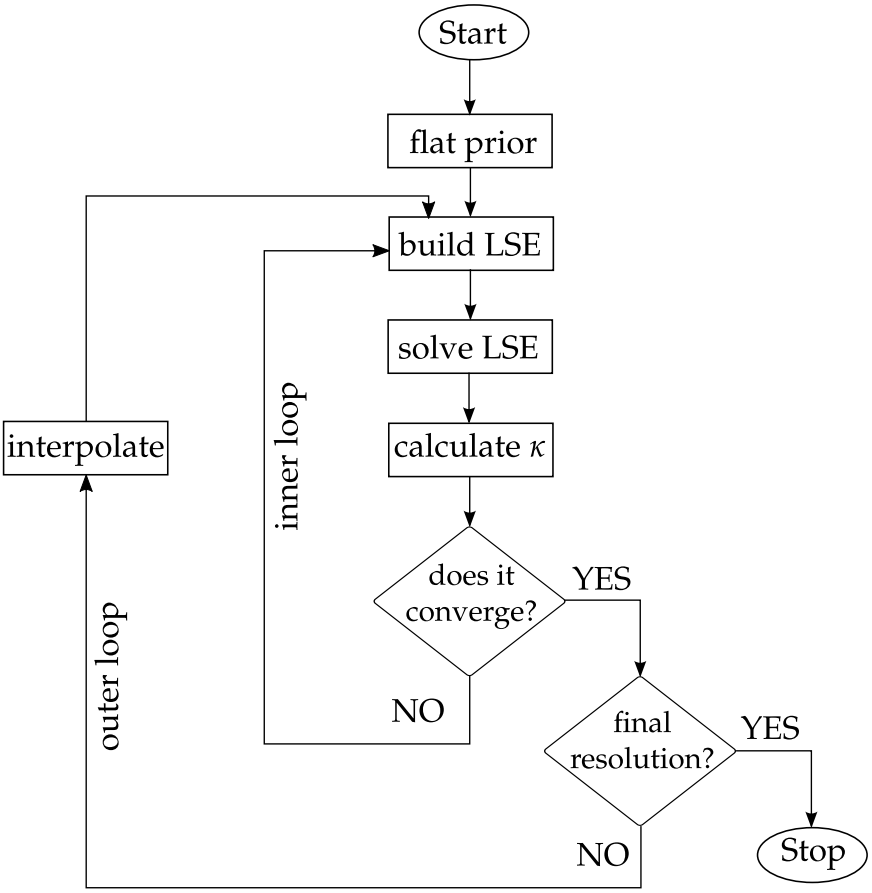

FIGURE 3.10: Workflow diagram depicting the SAWLENS2 iteration scheme

chapter, but falls short in cases of masking or significant over- or underdense regions in the background galaxy distribution. For these cases, other `field node builders`, which add additional nodes if the local number density of galaxies around a given node exceeds a predefined threshold, are supplied.

If strong lensing features like multiple images are used, the positions of these images form another set of reconstruction nodes that is incorporated in the mesh-free support structure. Therefore, not all nodes carry the same kind of data, but the lensing potential is reconstructed on the combined set of nodes.

Iteration scheme

In order to keep the contamination by shape noise in the shear measurements low and to treat non-linearities in the log-likelihood terms, SAWLENS2 implements a two-level iteration scheme depicted in Figure 3.10.

The outer loop gradually increases the resolution of the reconstruction by adding nodes to the shear-based mesh-free domain, while at the same time reducing the number of nearest neighbours considered in the shear averaging and covariance computation. Starting out with a relatively high number of nearest neighbours (say 40), the shape noise induced by the intrinsic

ellipticities of the sheared galaxies is efficiently suppressed (Bradač et al., 2005). But since the covariance matrix is required to be numerically invertible, this necessitates a comparatively low number of support nodes as otherwise the overlap between nodes becomes too large and consequently the covariance matrix too full to be reliably inverted. In the first iteration, the reconstruction is regularized against a flat prior of zero convergence throughout the field of view. This prior enters the reconstruction through the regularisation term and through the terms in the log-likelihoods in Eqs. (3.26, 3.38), that are kept fixed to ensure linearity of the responses with respect to ψ.

The inner loop repeats the reconstruction on the same domain, each time updating the prior to the convergence map obtained from the previous iteration. Once this process converges[16] and the difference in the reconstructed convergence maps between iteration falls below a predefined threshold, the code exits the inner loop, interpolates the reconstructed convergence map to the new mesh-free domain (containing more nodes) and thus starts the next iteration of the outer loop until the final resolution is reached.

With each iteration in the outer loop, the number of nearest neighbours is decreased, which increases the shape noise contribution. Since the reconstruction is however regularized against the convergence map recovered at lower resolution, over-fitting of shape noise patterns continues to be suppressed. On the other hand, the reduced number of nearest neighbours means a reduced shear averaging area per node, resulting in a higher sensitivity to small scale changes in the lensing potential.
Note that the explicit choice of number of support nodes, number of nearest neighbours, regularisation strength and convergence threshold makes a certain amount of arbitrariness and tuning unavoidable.[17]

Implemented X-ray constraints

As the deprojection of the X-ray surface brightness and the subsequent projection of the Newtonian potential are here carried out under the assumption of spherical symmetry, this only constraines the azimuthally averaged radial profile of the lensing potential. There is therefore no a priori natural support domain and thus no set of reconstruction nodes is created solely to represent X-ray data in this case. This changes if that assumption is relaxed, for example to ellipsoidal symmetry, allowing to produce a two-dimensional estimate of ψ. For such situations, we provide a `field node builder`, sampling support nodes from the projected potential map, which in turn is likely based on the pixels of the original observation. But for the present case of spherically symmetric deprojections, such an approach would induce systematic errors. Suppose a cluster that is clearly elongated in a direction perpendicular to the

[16] It is unfortunate that a description of the iterative reconstruction of the lensing convergence almost necessarily requires the verb "converge"

[17] as an experience-based rule of thumb, the number of nodes times the number of nearest neighbours should not significantly exceed the total number of galaxies in the shear catalogue and the regularisation strength should be of the order of the diagonal elements in the inverse data covariance matrices

line of sight. The lensing terms in the combined log-likelihood do not suppose any symmetry, so they pick up on the asymmetric morphology and steer the reconstruction accordingly. If however on each node that is at the same radius, the X-ray term enforces the same value of $\bar{\psi}_\mathrm{X}$ this then leads to conflicts and overstates the importance of the X-ray constraint as it would demand strict spherical symmetry in conflict with the lensing data. We therefore choose the operator $\mathcal{P}$ in Eq. (3.41) such that

$$P_{ij} = \begin{cases} N_i^{-1} & \text{if node } j \text{ falls in bin } i \\ 0 & \text{else,} \end{cases} \tag{3.77}$$

where N_i is the number of nodes that lie within the radial bin i. This way, the log-likelihood in Eq. (3.41) ensures that the reconstructed potential values ψ on (azimuthal) average follow the profile $\bar{\psi}_\mathrm{X}$ while still leaving enough freedom for local deviations from spherical symmetry required by lensing constraints, for example due to significant substructure or an overall non-spherical cluster morphology.

3.7 Analysis and error estimation

Even though the presented framework employs approximately linear responses to the model parameters throughout, the free-form character of the method, combined with the regularisation priors, makes standard measures for goodness-of-fit like the reduced χ^2 non-trivial to compute (Bishop, 2006) and the fact that several log-likelihood terms are only *approximately* quadratic in ψ renders the usefulness of such measures questionable (Andrae, Schulze-Hartung, and Melchior, 2010).

As an alternative way to assess the quality of the reconstruction we choose normalised data residuals. After the reconstruction we compute the residual vector

$$\boldsymbol{r} = \boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\psi}), \tag{3.78}$$

for the combined data set $\boldsymbol{d}$ and the corresponding combined response $\boldsymbol{m}(\boldsymbol{\psi})$. In case of uncorrelated data, we would now divide each entry of the residual vector by the standard deviation of the corresponding data point. Since all our data sets are however correlated we have to first transform the residual vector into the eigenbasis of the combined inverse data covariance, in which the latter is diagonal

$$\mathcal{T}^\mathrm{T}\mathcal{C}^{-1}\mathcal{T} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N), \tag{3.79}$$

where N is the total number of data points, λ_i are the eigenvalues of the combined inverse data covariance and $\mathcal{T}$ is the corresponding basis change operator. The normalised residuals are then

$$r_{\mathrm{n},i} = \left(\mathcal{T}^\mathrm{T}\boldsymbol{r}\right)_i \cdot \sqrt{\lambda_i}. \tag{3.80}$$

If the assumptions going into the reconstruction are valid and if the regularisation is set properly, the normalised residuals should follow a Gaussian distribution with zero mean and standard deviation of one. If their distribution is significantly non-Gaussian, the assumed likelihood term was a bad choice. If the distribution is Gaussian, but significantly narrower, the reconstruction is likely under-regularised, i.e. the reconstruction would be overfitting the data. Conversely, if the distribution is much wider, the regularisation may be too strong.

Apart from assessing the quality of the reconstruction, we also need to quantify the errors of reconstructed quantities, especially the convergence, as it directly relates to the projected density distribution. Several ways to estimate errors of a SAWLENS2 reconstruction are possible and I will present two[18].

Since the reconstruction is entirely based on Gaussian likelihoods, all information on errors on the parameters ψ_i and derived quantities can in principle be obtained from the Fisher matrix

$$F_{ij} \equiv \frac{\partial^2 \chi}{\partial \psi_i \partial \psi_j} = B_{ij} \tag{3.81}$$

where we identified the coefficient matrix from Eq. (3.7). By construction, its inverse is the covariance matrix of the reconstructed ψ_i. To get an error estimate for the convergence we therefore only have to transform the inverse Fisher matrix

$$\mathcal{C}_\kappa = \mathcal{K}^{\mathrm{T}} \mathcal{F}^{-1} \mathcal{K}, \tag{3.82}$$

using the FD operator for the convergence $\mathcal{K}$. The diagonal elements of this matrix are then the variance of the reconstructed convergence at the respective nodes. The downside of this method of error estimation is, that it requires the explicit inversion of the coefficient matrix, unlike in the actual solving of the LSE. Since the coefficient matrix is rather dense, numerical inversion is not guaranteed to be stable, and some form of matrix conditioning, for example Tikhonov regularisation (Tikhonov et al., 1995), is often needed to make it feasible at all. We nevertheless compute this covariance estimate in SAWLENS2 and present the result in the next section.

The standard way of error estimation in SAWLENS2, however is resampling, specifically bootstrapping (Andrae, 2010; Merten, 2016). In this mode, SAWLENS2 creates a number of realisation of the reconstruction by resampling (with replacement) the entire shear catalogue and randomly sampling the source redshift of strong lensing constraints from their confidence range. The X-ray based lensing potential profile is also bootstrapped, by resampling the Poisson realisations created for the covariance estimation and averaging over the resampled set instead of the original to create the mean projected profile, but keeping the original covariance matrix.

Since the shear catalogue gets resampled for each bootstrap realisation and as there is an element of randomness in the creation of the mesh-free domain from it anyway, all these realisations lie on slightly different domains.

[18] Another error estimation for lensing-only reconstructions can be found in Majer (2014)

In order to create an estimate of the errors on the default reconstruction, we therefore interpolate the bootstrapped convergence maps onto the domain of the default map, before computing averages and standard deviation. We have to point out, that, most likely due to numerical instabilities along the way, a small fraction of bootstrap realisations returns entirely unphysical results, like convergence values several orders of magnitudes above the weak lensing regime in a reconstruction without strong lensing constraints. We identify and discard such obviously corrupted cases before the error estimation.

Chapter 4

Application

In the following, I will present applications of the reconstruction framework described in Section 3.2 to simulated clusters, in order to detail the intended usage, features and potential advantages and pitfalls of our method of combining lensing and X-ray data in a joint reconstruction. The first, most basic use case will be a simple, spherically symmetric NFW halo, the second a realistic cluster taken from a hydrodynamical simulation of large scale structure.
The results in this chapter are also presented in Huber et al. (2018).

4.1 Simple mock

For the most basic test of our method, we chose a case where all assumptions are true by construction, i.e. a NFW halo with a polytropic ICM in perfect hydrostatic equilibrium. We set the virial mass of the halo to $M_{\mathrm{vir}} = 5.0 \cdot 10^{14} M_{\odot} h^{-1}$ at redshift $z_{\mathrm{l}} = 0.3$, which according to the concentration-mass-relation

$$c = 6 \left(\frac{M_{\mathrm{vir}}}{10^{14} M_{\odot} h^{-1}} \right)^{-1/5}, \tag{4.1}$$

(Seljak, 2000; Klypin et al., 2001) results in a concentration of roughly $c \approx 4.71$ and a virial radius of

$$R_{\mathrm{vir}} = \left(\frac{M_{\mathrm{vir}}}{\frac{4\pi}{3} \Delta_{\mathrm{c}} \rho_{\mathrm{crit}}(z)} \right)^{1/3} \approx 1.71\,\mathrm{Mpc} \tag{4.2}$$

(Peebles, 1980), where we used the critical overdensity $\Delta_{\mathrm{c}} = 200$. The lensing properties of NFW halos are analytically known (Golse and Kneib, 2002), so we can easily create[1] a mock shear catalogue with sources at $z_{\mathrm{s}} = 1.0$ and a source number density of $n_{\mathrm{gal}} \approx 20\,\mathrm{arcmin}^{-2}$. Setting the field of view to 21 arcmin $\times$ 21 arcmin, this results in roughly 6900 randomly distributed galaxies with shears to which we add Gaussian shape noise with $\sigma = 0.3$. We set the gas mass fraction to $f_{\mathrm{gas}} = 0.2$ and follow the formalism of Komatsu and Seljak (2001) to model the ICM in hydrostatic equilibrium and with a polytropic stratification with index $\gamma = 1.13$. We create a surface brightness

[1] Part of this was done using the libastro code library developed by the cosmology group at ITA Heidelberg.

Halo	Lensing	X-ray
$M_{\mathrm{vir}} = 5.0 \cdot 10^{14} M_{\odot} h^{-1}$	FoV$= (21\,\mathrm{arcmin})^2$	FoV$= (21\,\mathrm{arcmin})^2$
$R_{\mathrm{vir}} = 1.71\,\mathrm{Mpc}$	$z_{\mathrm{s}} = 1.0$	$\hbar\bar{\omega} = 1.5\,\mathrm{keV}$
$c = 4.71$	$n_{\mathrm{gal}} \approx 20\,\mathrm{arcmin}^{-2}$	$t_{\mathrm{exp}} = 3000\,\mathrm{s}$
$z_{\mathrm{l}} = 0.3$	$\sigma = 0.3$	$A_{\mathrm{eff}} = 500\,\mathrm{cm}^2$
$f_{\mathrm{gas}} = 0.2$		
$\gamma = 1.13$		

TABLE 4.1: Defining parameters for the NFW halo and the mock observations of it;

map with the same field of view and assuming a mean photon energy of 1.5 keV, an exposure time of 3000 s and a detector with an effective area of 500 cm^2. All defining properties of the halo and the mock observations are collected in Table 4.1 and for completeness, all relevant equations for the creation of mock observations of a NFW can be found in Appendix B.

We add Poissonian noise to the resulting X-ray surface brightness before creating 20 000 Poisson realisations of the surface brightness profile in 25 linearly spaced radial bins, as described in Section 3.6.1. We deproject these realisations with the RL algorithm and, using the correct polytropic index and temperature of roughly 8.1 keV, determine the Newtonian potential profile. Based on the realisations of $\Phi_{\chi}(r)$, we estimate the signal-to-noise ratio of the enclosed mass as a function of radius, again as discussed in Section 3.6.1 and depicted in Figure 3.8. Additionally we compute the Laplacian of the potential as as a proxy for the total density, illustrated in Figure A.2. We now discard all bins that fail our sanity checks, i.e. where the Laplacian is cleary negative and where the signal-to-noise ratio falls below a threshold value of 3. The thus truncated potential profile is subsequently projected along the line of sight and using the analytic continuation according to Section 3.5.5 up to a cut-off radius of three times the truncated data radius. The resulting lensing potential profile is compared to the analytically known true profile in Figure A.3. Even without knowing the true lensing potential an appropriate cut-off radius for the analytical continuation can be found by gradually increasing it from the truncated data radius. If the cut-off radius is too small, the projection artefacts manifest themselves in deviations in the radial curvature of the potential at increasing radii (with increasing cut-off, see Fig. 3.6). These curvature deviations lead to ring-like structures on the reconstructed convergence map, which are easily spotted as unphysical. This of course doesn't pin down the proper cut-off radius entirely, but gives a lower limit. The realisations of ψ_{χ} are Gaussian distributed as shown in Figure A.4, so the assumption of a Gaussian likelihood is valid. We scale all realisations to a reference redshift of $z_{\mathrm{inf}} = 20\,000$ and use them to compute the inverse covariance matrix of the projected potential profile.

Nodes (nominal)	300	500	800
Nearest neighbours	33	21	12
Regularisation strength	200	400	600

TABLE 4.2: List of parameters defining the outer loop iterations for the SAWLENS2 recosntruction of a NFW halo; The algorithm has some leeway when subsampling the shear catalogue to set up the mesh-free domain, so the actual numbers of nodes deviates slightly from the nominal one. The regularisation strength is set to be the same for both, the regularisation against the convergence and the one against the shear (see Section 3.4)

With this, all ingredients are in place and we run SAWLENS2, once only using the shear data (designated as W from here on) and once with both shear and X-ray based constraints (W+X hereafter). In both cases, the mesh-free support is defined by subsampling the shear catalogue as described in Section 3.6.2 and we choose three resolution levels, with 300, 500 and 800 nodes. All parameters of the reconstruction are collected in Table 4.2 and as an example the distribution of nodes in the highest resolution is shown in Figure A.1.

In order to estimate errors on the resulting convergence maps, we run 500 bootstrap realisations of these reconstructions each. Figures A.9 and A.10 show the results of the default reconstruction with relative residuals based on the Fisher error estimation (see Fig. A.8), whereas figures A.5 and A.6 show the mean result of the bootstrapped reconstructions with relative residuals based on the bootstrapped errors (see Fig. A.7). The convergence profiles based on the default reconstructions and the mean of the bootstrap realisations are shown in figures A.11 and A.12 respectively and Figure 4.1 depicts the profiles of the default reconstruction with the bootstrapped error bars.

To assess the quality of the reconstruction we show the normalised residuals (compare Section 3.7) in Figure A.13. We can summarise the results of this run as follows.

- The central convergence peak is significantly more pronounced in the combined reconstruction than in the shear-only case.
- The error estimate based on the Fisher matrix clearly and persistenty (in both runs) shows numerical artefacts not related to any structure or noise pattern in the data.
- The bootstrap-based error estimates for both cases are very similar with very slightly increased errors for the combined case.
- The bootstrap-based error estimates increase towards the central peak, as expected, since picking a slightly different node position has the largest effect where the gradient of the convergence is large.

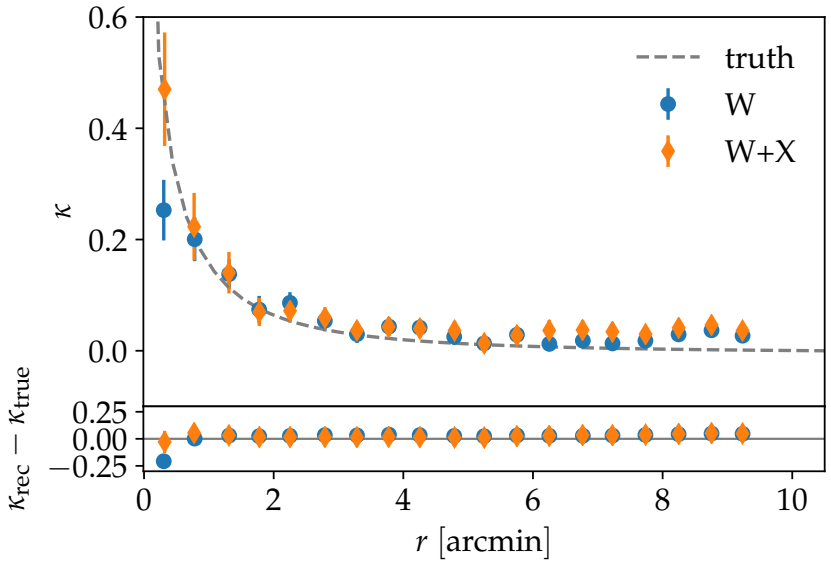

FIGURE 4.1: Reconstructed convergence profiles for a NFW mock cluster, with error bars inferred from averaging over bootstrapped samples;

- The convergence profile based on the bootstrapped reconstruction indicates a slight bias towards higher convergence values in the lensing reconstruction and the outskirts of the combined. The origin of this is not entirely clear, but could lie in under-regularisation.
- the relative residuum $(\kappa_{\rm rec} - \kappa_{\rm true})/\sigma_{\kappa,\rm BS}$ with respect to the bootstrapped error estimation shows better agreement in the combined case and indicates a fairly unbiased reconstruction.
- The normalised data residuals are in reasonable agreement with expectations, even though a slight tendency towards over-fitting is visible.

The addition of an X-ray based constraint on the radial profile of ψ can significantly improve the reconstruction, at least in cases where the assumptions going into it hold true and where otherwise only shear data of average quality is available.

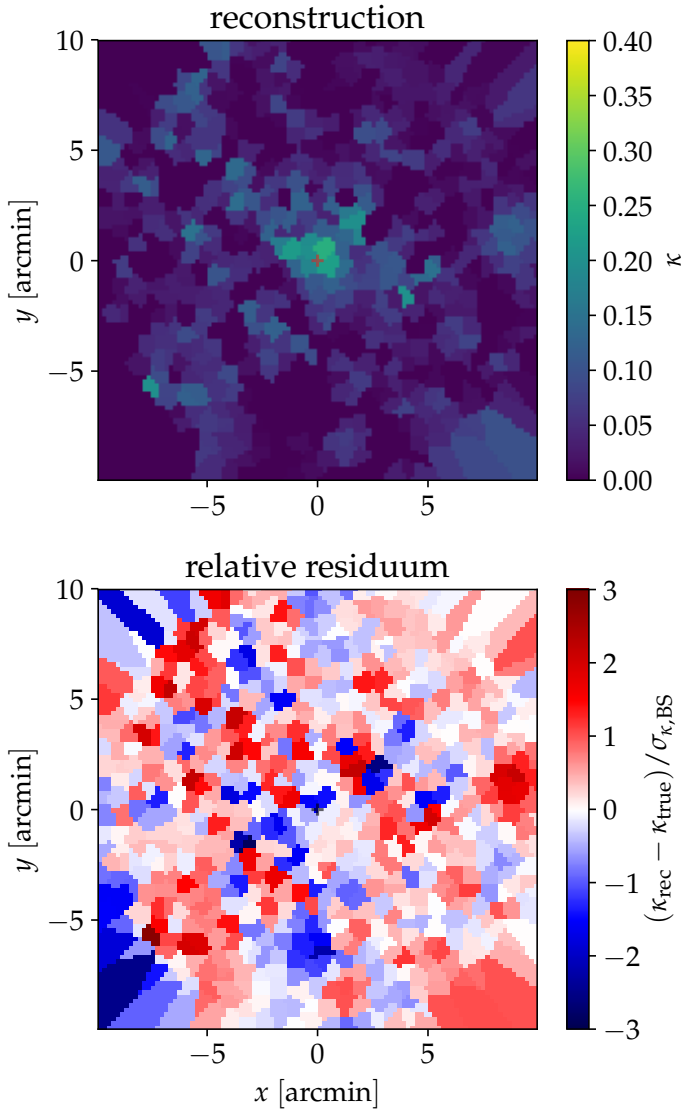

FIGURE 4.2: (*top*) Reconstructed convergence map for a NFW mock cluster, based on weak lensing alone; (*bottom*) relative residuum, compared to error estimate based on bootstrapping;

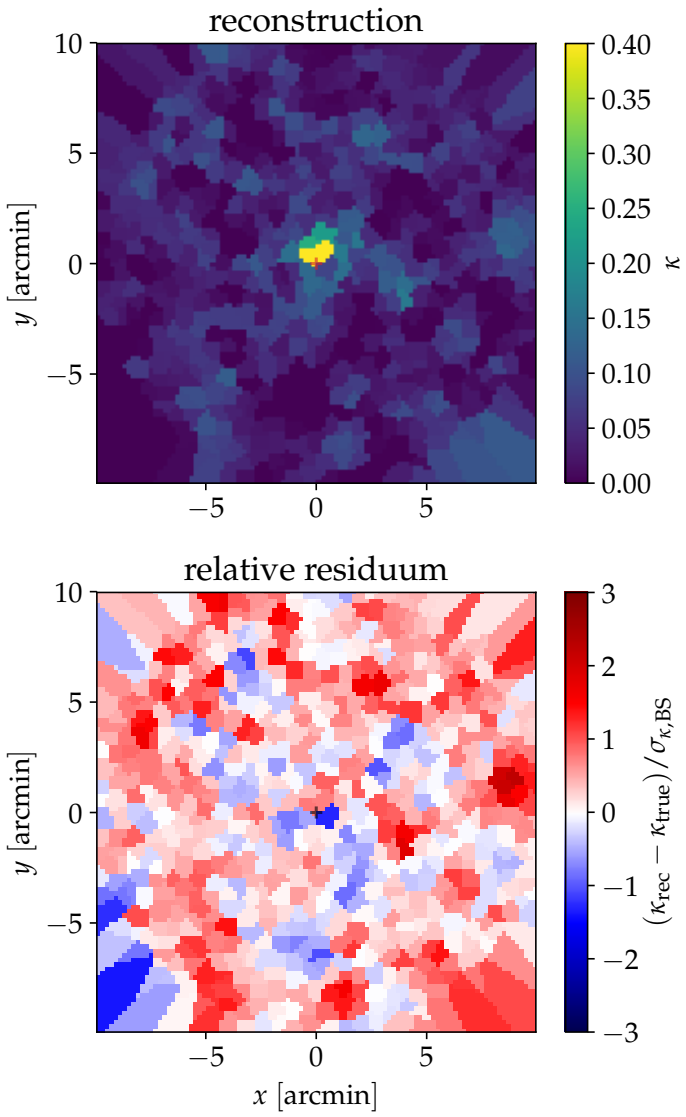

FIGURE 4.3: (*top*) Reconstructed convergence map for a NFW mock cluster, based on weak lensing and X-ray data; (*bottom*) relative residuum, compared to error estimate based on bootstrapping;

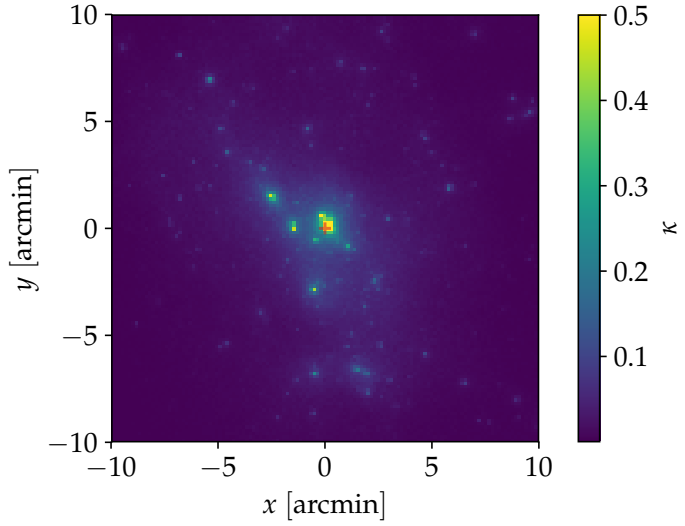

FIGURE 4.4: Convergence map of the realistic simulated cluster; It clearly features substructure and deviates from spherical symmetry.

4.2 Realistic simulation

To investigate the performance of the reconstruction framework when faced with more realistic cluster morphologies, we turn to a hydrodynamical simulation of a cluster embedded in the cosmic web. We make use of the public web interface[2] first presented in Ragagnin et al. (2017) to select a massive cluster from Box2b/hr of the *Magneticum*[3] (Dolag et al. in prep.; Hirschmann et al., 2014) simulation suite. The simulations, as well as our subsequent treatment of them in the context of this work, adopt a WMAP7 cosmology (Komatsu et al., 2011) and the particular box follows $2 \cdot 2880^3$ particles in a volume of $(640\,\mathrm{Mpc}/h)^3$, using the N-body/SPH code `P-Gadget3` (Springel, 2005; Beck et al., 2016) and it includes the full variety of non-gravitational effects, like radiative cooling, heating, star formation and AGN feedback.

The cluster used here has a mass of $M_{500} = 5.54 \cdot 10^{14} M_\odot/h$ at a redshift of $z = 0.25$. Its convergence map is shown in Figure 4.4 and Table 4.3 collects some of its properties. We select this particular cluster, because it displays a structured and non-spherical morphology while being among the more massive halos in the box, which allows for pristine enough mock observations.

We employ the tools PHOX (Biffi et al., 2012) and SIXTE[4] (Schmid et al., 2010), which are both implemented in the aforementioned interface, to generate synthetic X-ray observations with $4 \cdot 10^4$ s exposure with XMM-Newton's EPIC-pn instrument and use SMAC (Dolag et al., 2005), to obtain a total surface

[2] `https://c2papcosmosim.uc.lrz.de`
[3] `http://www.magneticum.org/`
[4] `http://www.sternwarte.uni-erlangen.de/research/sixte/`

Halo	Lensing
$M_{500} = 5.54 \cdot 10^{14} M_\odot h^{-1}$	FoV$= (21.3\,\mathrm{arcmin})^2$
$R_{500} = 1.14\,\mathrm{Mpc}$	$z_\mathrm{s} = 1.0$
$z_\mathrm{l} = 0.25$	$n_\mathrm{gal} \approx 25\,\mathrm{arcmin}^{-2}$
$T_{500} = 5.98\,\mathrm{keV}$	$\sigma = 0.3$
$f_\mathrm{gas} = 0.13$	

TABLE 4.3: Defining parameters for the realistic simulated cluster and the mock lensing observations of it;

density map. Based on the latter we compute the convergence map and from that the complex shear via the relation

$$\gamma(\boldsymbol{\theta}) = \frac{1}{\pi} \int \mathrm{d}^2\theta' D(\boldsymbol{\theta} - \boldsymbol{\theta}')\kappa(\boldsymbol{\theta}'), \tag{4.3}$$

with the convolution kernel

$$D(\boldsymbol{\theta}) = \frac{-1}{(\theta_1 - \mathrm{i}\theta_2)^2}. \tag{4.4}$$

With this we can generate a map of the reduced shear g and sample it at a source density of $n_\mathrm{gal} \approx 25\,\mathrm{arcmin}^{-2}$ to generate a catalogue of weakly lensed sources at redshift $z_\mathrm{s} = 1$, adding again Gaussian noise with $\sigma = 0.3$.

We additionally create a map of ICM temperature, to estimate the sound speed, and a map of the tSZ signal of the cluster. Deprojecting the latter, as well as the count map based on the synthetic XMM observations depicted in Figure A.14, we can estimate the polytropic index of the ICM. To do so, we plot the electron pressure obtained from deprojecting the Compton-y profile against the X-ray emissivity, realising that, if the assumptions in Section 3.5.2 hold, both are related by

$$P(r) \propto j_\mathrm{X}^{2\gamma/(3+\gamma)}(r). \tag{4.5}$$

Figure 4.5 shows the pressure versus the emissivity in arbitrary units and the fit based on Eq. (4.5), excluding areas, where the simple powerlaw relation seems broken and the assumptions likely don't hold. The best fitting polytropic index according to this then is $\gamma = 1.10$.

As in the previous example, we again create 20 000 Poisson realisations of the count profile which are deprojected, converted to realisations of the Newtonian potential and subjected to the same sanity checks. Figures A.15 and A.16 show the signal-to-noise ratio of the enclosed mass and the Laplacian of the potential respectively, as well as the radial bins excluded from the further analysis. The thus truncated potential profiles are then projected,

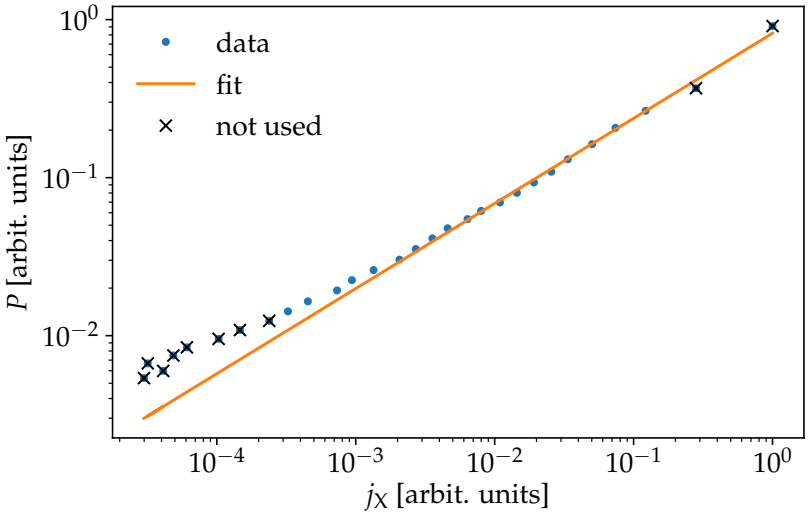

FIGURE 4.5: The pressure versus the emissivity in the realistic cluster. The data displays a broken powerlaw behaviour indicating that the simple assumptions going into our method are not valid throughout. We restrict the analysis and the fit to obtain the polytropic index to the radial region where the P-j_X follows a straight line in double-logarithmic scaling;

again employing the analytic continuation scheme described above up to a cut-off radius of four times the truncated data radius. The resulting ψ potential realisations are used to estimate the corresponding inverse covariance matrix and their mean profile is compared to the actual lensing potential of the simulated cluster in Figure 4.6. The curves agree well, up to a radius of roughly 5 arcmin, where the X-ray-based estimate starts to deviate, however still well within the error margins. We chose the cut-off radius based on ballpark estimates of the virial radius from the surface brightness data and experience from Section 3.5.5. The error bars in Figure 4.6 are noticeably larger than in the NFW case above, and one reason for this is the choice of cut-off radius. The further out the extrapolation is taken, the more noisy features, even in the truncated potential profile, influence the continuation and thus the projection and consequently the variance throughout the realisations increases. A balance between extrapolating far enough to counter artefacts while still maintaining constraining power needs to be found, but so far rough estimates have proven to work sufficiently well. The realisations again are compatible with a Gaussian likelihood, as shown in Figure A.17.

The following SAWLENS2 reconstructions, as well as the 500 bootstraps, are set up using the parameters collected in Table 4.4. Again, we perform one reconstruction, only using the shear catalogue and one, combining shear and X-ray data. The resulting convergence maps are shown in figures 4.8 and 4.9, together with their relative residua based on bootstrapping, whereas the mean results of the bootstrapping are depicted in figures A.18 and A.19. Figure A.24 shows the convergence profiles based on the default reconstructions,

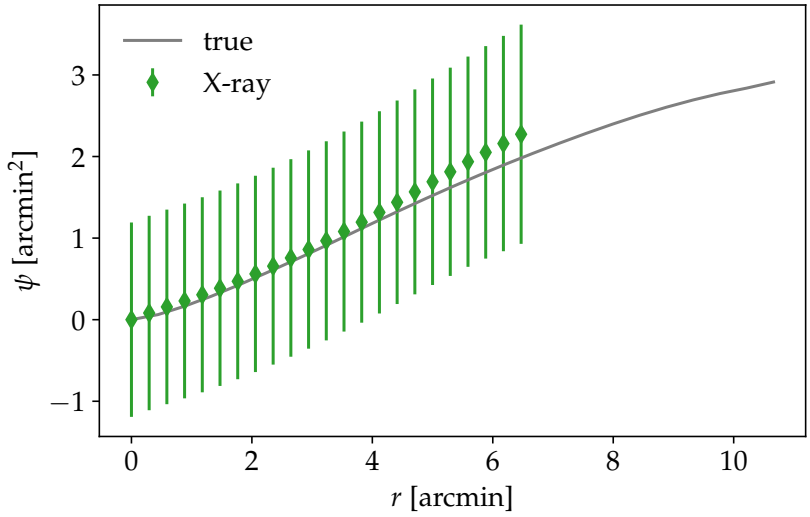

FIGURE 4.6: The lensing potential of a realistic cluster, obtained from projecting the X-ray based estimate of the Newtonian potential;

Nodes (nominal)	300	550	800
Nearest neighbours	40	22	15
Regularisation strength	200	400	600

TABLE 4.4: List of parameters defining the outer loop iterations for the SAWLENS2 recosntruction of the realistic cluster; The regularisation strength is set to be the same for both, the regularisation against the convergence and the one against the shear (see Section 3.4)

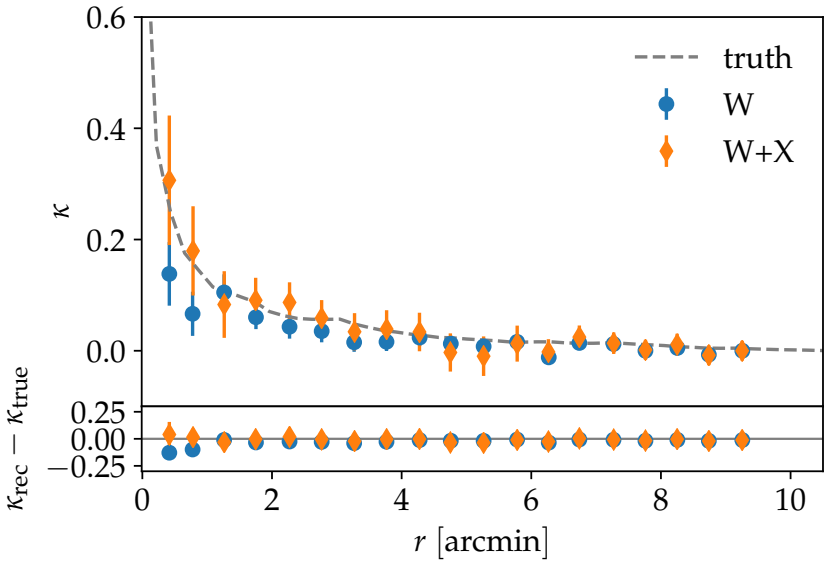

FIGURE 4.7: Reconstructed convergence profiles for a realistic cluster, with error bars inferred from averaging over bootstrapped samples; curves shifted to coincide at outermost bin to correct slight bias to higher convergence values (see text);

Figure A.25 those based on the bootstraps and Figure 4.7 depicts the profiles of the default reconstruction with the bootstrapped error bars. The estimated error maps from Fisher analysis and bootstrapping are shown in figures A.21 andA.20 respectively and the normalised residuals are presented in Figure A.27.

As in the previous use case, the combined reconstruction results in a more pronounced central convergence peak and the radial convergence profile nicely traces the true one, especially in the bootstrapped mean. We do however note, that the combined reconstruction is noticeably biased high in convergence, as can be seen in the relative residuum map in the lower plot of Figure 4.9 and in Figures A.26 and A.28. Experiments with increased and decreased cut-off radii in the projection step for the X-ray constraint show that the bias is not related to deviations in the projected potential curvature alone. The scale of the lensing potential is, as Figure 4.6 shows, well reproduced and as the previous section showed, the combination of shear- and X-ray-based constraints in SAWLENS2 does not generally introduce a bias in the reconstructed convergence. Further investigation reveals that this bias originates in the combination of two factors. The recovered profile of the lensing potential based on X-ray data, $\bar{\psi}_X$, deviates slightly from the truth (see figure 4.6) and the covariance matrix $\mathcal{C}_X$ is very dense, effectively correlating every radial bin with every other as can be seen from the Pearson correlation matrix depicted in Figure A.29. This way, slight overestimations of the radial curvature in the outer bins lead to slight additive biases in the convergence of the joint reconstruction. Since this is a purely additive bias, it can be mitigated relatively easily. It

highlights however once more the need for careful treatment of artefacts in the projection step, as they can alter the results across the whole domain. The radial profiles in figure 4.7 are adjusted for this bias.

In order to quantify how the reconstruction deals with the obviously triaxial morphology of the cluster, we calculate the second moment tensor

$$E = \sum_n \kappa_n \Theta \left(\kappa_n - \kappa_{\text{thresh}} \right) \boldsymbol{\theta}_n \otimes \boldsymbol{\theta}_n \tag{4.6}$$

for the convergence map, where the sum runs over all nodes in the reconstruction and the Heaviside function $\Theta(\kappa_n - \kappa_{\text{thresh}})$ ensures that only nodes with convergence above a certain threshold are taken into consideration. The eigenvectors of this tensor align with the major axes of the halo, while the ratio of its eigenvalues reflects the square major axes ratio.

The corresponding eigenvectors are shown in Figure 4.10, scaled to reflect the reconstructed eigenvalue ratio. The X-ray-based information does not add information on the morphology as it is based on spherical symmetry to create an estimate of the radial profile of the lensing potential. But since the reconstruction is only required to follow this profile on azimuthal average, it also does not impose any morphology. In cases where the noise in the reconstruction is reduced by the addition of X-ray data, the constraints on the morphology may be improved. In the present case, where we used the median reconstructed convergence as a threshold for Eq. (4.6), the reconstructions only poorly reflect the axis ratio of the true map, but the combined reconstruction matches the directions of the major axes significantly better than in the shear-only case.

This is however not a particularly stable result. Changes in the iteration and regularisation settings may alter the inferred second moment tensor without significantly changing the normalised data residuals or other reconstruction results. It therefore remains doubtful if the addition of symmetrised X-ray constraints can improve the inference of the morphology of a real cluster. We do however expect, that X-ray constraints using deprojections with spheroidal symmetry are bound to improve this measure, as Figure A.14 shows a clear elongation of the surface brightness, paralleling that of the underlying mass distribution.

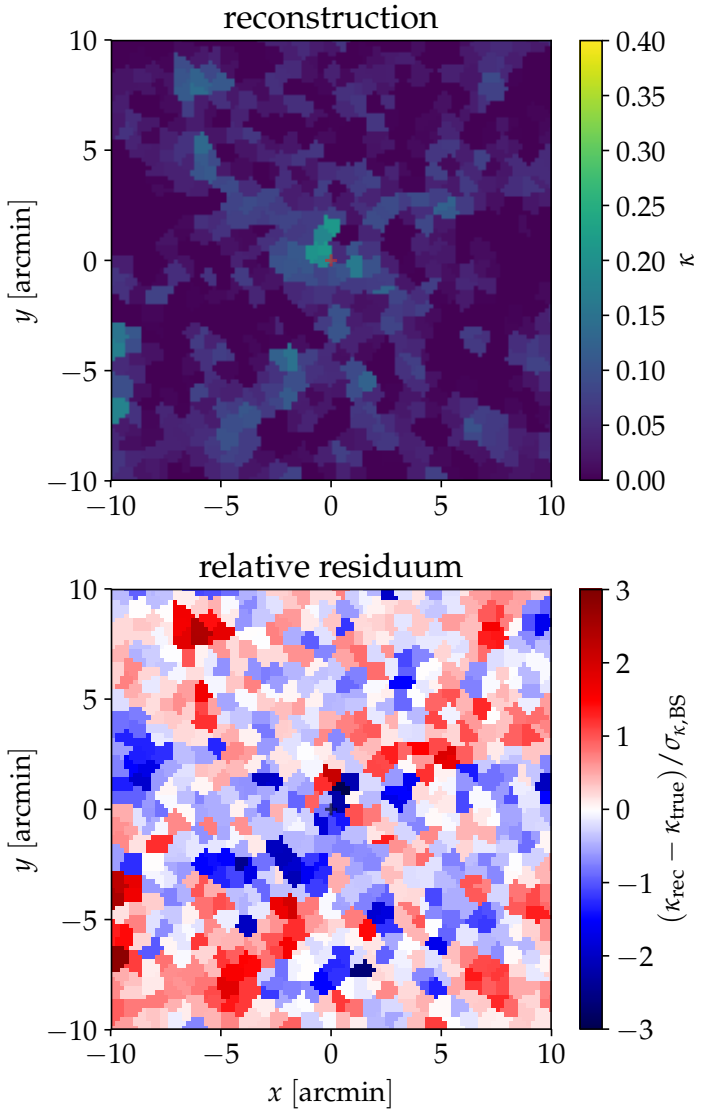

FIGURE 4.8: (*top*) Reconstructed convergence map for a realistic cluster, based on weak lensing alone; (*bottom*) relative residuum, compared to error estimate based on bootstrapping;

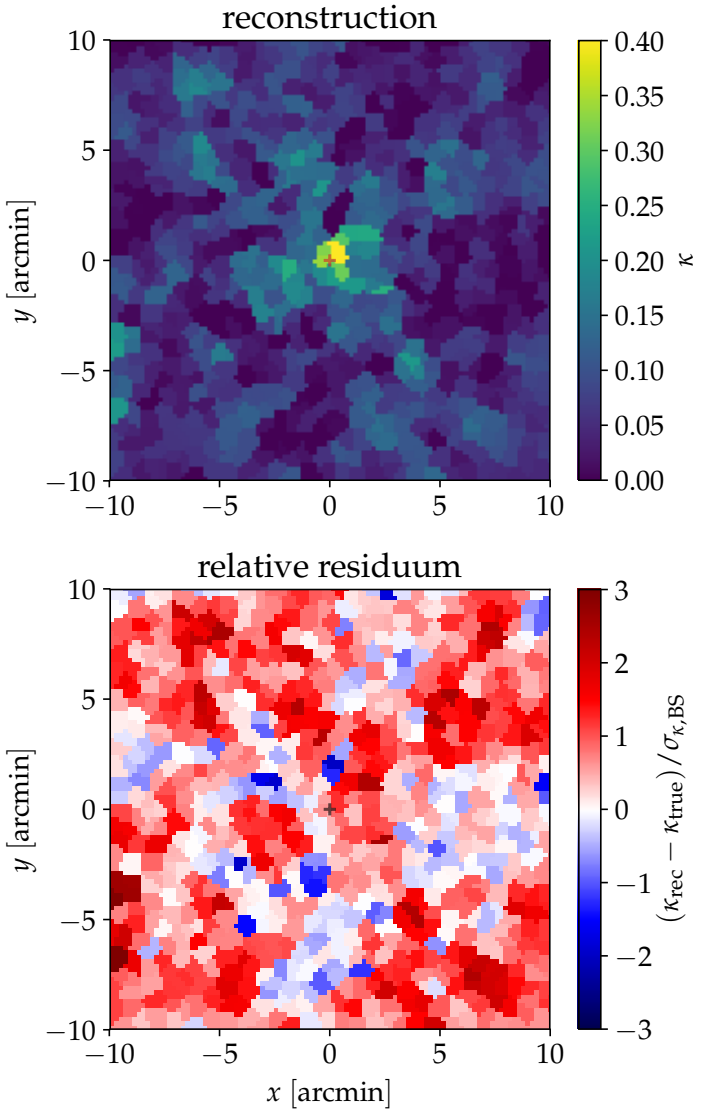

FIGURE 4.9: (*top*) Reconstructed convergence map for a realistic cluster, based on weak lensing and X-ray data; (*bottom*) relative residuum, compared to error estimate based on bootstrapping;

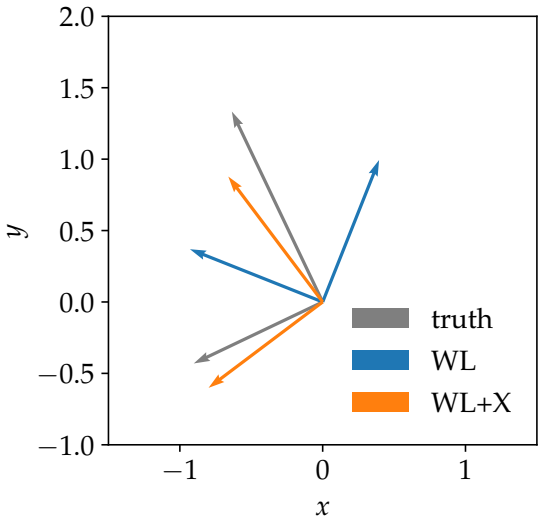

FIGURE 4.10: The eigenvectors of the 2nd moment tensor for the reconstructions of a realistic cluster, compared to the ones based on the true convergence map; The combined case (orange) shows better agreement with true orientations of the cluster than the shear-only case (blue), but both fail to recover the axis ratio.

Chapter 5

Conclusion

5.1 Summary

The cosmic large scale structure in general and clusters of galaxies in particular are used in a plethora of ways to further our understanding of the Universe as a whole and of the structures we find in it. In order to make the best possible use of the rich observational data available for clusters, we require means to understand individual clusters in detail and to test the assumptions we make when modelling them in bulk.

After establishing the wider context and highlighting some of the many features and use cases of galaxy clusters in chapters 1 and 2, I discussed in chapter 3, first in general, how methods to reconstruct the mass distribution of clusters can be classified and what use case calls for which type of method. I then continued to introduce the mesh-free and free-form framework we settled on and its most recent implementation, SaWLens2. I presented how SaWLens2, initially designed to combine strong and weak lensing data, can make use of additional constraints based on gaseous and kinematic tracers of the underlying gravitational potential. I introduced the classic Richardson-Lucy algorithm as a means to deproject such observations in order to derive the Newtonian potential of a cluster and showed how this can be done maintaining the appropriate normalisation and dimensionality. Subsequently, the issue of hitherto unnoticed numerical artefacts in the process of reprojecting the Newtonian potential presented itself and I discussed a scheme to amend these problems in a data-driven way. Finally, I suggested a measure to assess the quality of reconstructions in application cases and demonstrated how we can estimate errors in a SaWLens2 reconstruction.

In Chapter 4, I then moved to testing the reconstruction framework. I first showed how weak lensing and X-ray mock observations of a simple NFW halo can be combined in SaWLens2, to improve the accuracy of the reconstruction. After that, I tested the method in a more realistic, yet still controlled example, tackling mock observations of a realistic galaxy cluster from a state-of-the-art hydrodynamical simulation. We saw that, while many of the improvements first seen with the NFW halo hold true, the addition of X-ray-based constraints, especially under the strong symmetry assumptions imposed by the present implementation of the Richardson-Lucy algorithm, is not an universal remedy

and does introduce complications and potential biases that need to be weighed against the undeniable advantages in a case by case fashion.

5.1.1 Assumptions

In the course of this thesis several limiting and simplifying assumptions have been made and it is worthwhile to collect them here in a concise list and point out how they affect this work.

We assume that the ΛCDM framework is an adequate description of the Universe. This entails the assumption that

- GR is the appropriate theory of gravity on all scales relevant here. Our method of joint analysis is therefore in its basic form unfit to test deviations from GR.
- particle dark matter (cold or warm) is the appropriate explanation for cosmological observations and dominates the gravitating matter in the Universe and particularly in galaxy clusters.

In the context of gravitational lensing, we limit ourselves to cases that can be sufficiently described in the single thin lens approximation and assume that intrinsic alignment is a negligible source of systematic error on scales relevant for cluster lensing.

In the context of using X-ray data as an additional constraint, we assume that

- the cluster at hand is in approximate hydrostatic equilibrium. This is needed to connect the gas structure to the total Newtonian potential.
- the ICM is polytropically stratified over a significant radial range.
- the ICM can be described as an ideal gas.

For our reconstruction method, we assume that Gaussian likelihoods are sufficient to describe the data and their uncertainties, and specifically we assume that the statistical uncertainties in X-ray observations of galaxy clusters are purely Poissonian, when we create the realisations necessary to estimate the covariance matrix of the projected, X-ray-based potential.

Some of these assumptions can be altered or replaced. For instance, parametrised deviations from GR (Zhao et al., 2009) can in principle be incorporated to either allow a consistent reconstruction assuming modified gravity or to test for such deviations. More involved lensing scenarios, where for example intrinsic alignment is taken into account can be implemented by altering the computation of the shear covariance matrix. Similarly, corrections for hydrostatic bias can be included. It is however doubtful that any of the assumptions in this section can be dropped entirely without altering the core concepts of the methods they are involved in.

5.2 Discussion and outlook

5.2.1 Scope of the framework

SAWLENS2 is a free-form method and as such not equipped for reconstructions of very high statistical precision. It cannot replace parametric models of gravitational lenses such as galaxy clusters, it can only complement them. To do so, it has to be able to constrain properties of a cluster that a parametric method does not capture. One very simple and perhaps educational example not mentioned so far is the issue of miscentering. A parametric model has to be centred on a specific point[1] in space or on the sky. This is often done by either choosing the BCG as a center or the peak of the observed surface brightness or tSZ map, but that does not necessarily have to be the center of the total mass distribution. In its lensing-only configuration, SAWLENS2 does not care about the position and layout of the cluster with respect to the coordinate frame, as long as it is in it, and therefore can readily be used to, at least approximately, locate the peak of the total projected mass distribution. In this way it can inform a statistically more precise parametric model. This small example already showcases that the addition of X-ray constraints does not necessarily always improve everything. While Chapter 4 showed that it makes the central convergence peak much more pronounced, the deprojection, due to the underlying symmetry assumptions, itself has to be centred first, somewhat defying the point of the exercise. The lesson is, that the usefulness of any specific probe as well as of any specific combination of probes depends on the goal in mind. The great advantage of SAWLENS2 in its current form is, that it is now equipped to handle essentially all these probes as well as all their combinations and therefore is a most versatile tool.

SAWLENS2, by now, is a conveniently fast method, taking only between seconds and few minutes for reconstructions of the type presented here[2], allowing to quickly perform several exploratory reconstructions of a cluster once the necessary preparations are done. For this, one has to set the hyper-parameters of the reconstruction, like the regularisation strength and the specifics of the iteration and shear averaging scheme, by hand however, which introduces a degree of arbitrariness to the reconstruction. Also the cut-off radius for the analytic continuation has to be set by hand. While there are reasonable experience-based guidelines how to choose these parameters, there is still considerable leeway and results may be affected by that, as the examples of reprojection and the second moment tensor in Section 4.2 illustrate. These parameters could of course in principle be optimised for, either beforehand by calibrating the method to a training set of simulations and mocks, or on the fly, by iteratively fixing hyper- or model parameters. The former case would require automatising the entire preparatory pipeline described in Section 3.6.1, the latter a measure of goodness of fit, that is significantly more sensitive

[1] even though the centering can of course also be parametrised

[2] on a regular laptop; bootstrapping of course still takes a while

to changes in any of the hyper-parameters than the normalised data residuals. Their quantitative assessment is complicated by numerical instabilities, which occasionally lead to a single extreme outlier, throwing off quantitative measures of statistical compatibility, like the Kolmogorov-Smirnoff or the D'Agostino-Pearsons normality test. The root of these instabilities lies in the inversion of the data covariance matrices, which in many cases can be poorly conditioned.

While the tools to connect non-lensing data to the lensing potential rely on sometimes debatable assumptions, they can, for the same reasons , be used to test these very assumptions. Using the RL machinery to deproject X-ray and tSZ data allows to test polytropic stratification, and when lensing and X-ray based estimates of the potential behave significantly differently, the reason may lie in hydrostatic bias.

To summarise, despite some shortcomings, the methods presented here offer means to investigate systematic uncertainties both for scaling relations and mass calibrations in large cosmological surveys as well as for studies of individual lens systems.

5.2.2 Future prospects

There are some more or less obvious avenues for this particular project to continue from here.

Perhaps the most obvious is application. While SAWLENS2 has been and continues to be used for lensing based reconstructions of galaxy clusters, the combination of lensing and X-ray (or other) constraints yet has to be taken to the real world. This thesis is intended as a guideline for such applications and a number of excellent, already available data sets, first and foremost maybe the CLASH survey of X-ray selected massive clusters (Postman et al., 2012), can readily be approached with the framework in its current state layed out here.

The inclusion of non-lensing constraints as presented here, relies on spherical symmetry. In Reblinsky (2000) and Majer et al. (2016) the Richardson-Lucy algorithm got extended to spheroidal symmetries. This complicates matters in the deprojection step as the inclination angle of the spheroid with respect to the line of sight has to be inferred in some way. The projection step and the artefact treatment it involves, however should not be affected by this generalisation. The actual incorporation of the estimated projected potential in SAWLENS2 can then happen on the map level instead of the azimuthal average, and the modular structure of the code framework readily allows to include such a constraint.

Chapter 3 already hints at the incorporation of source inversion techniques in SAWLENS2. Parametric and semi-parametric source inversion techniques have become important tools, both for cosmology, as in tests of dark matter via the measurement of line-of-sight structure in strong lensing by galaxies (Despali et al., 2018) or measurements of cosmic expansion via time delays, as

well as for studies of the high redshift source galaxies themselves. These studies often rely on the separability and sufficiency of a relatively well-behaved main lens and a simple external shear contribution, which is questionable for strong lenses in many environments but especially in galaxy clusters, as their extent and formation history allows for a variety of morphologies of varying complexity. Free-form reconstruction techniques on the other hand are ideally suited to constrain the potentially complex mass distributions in such environments. Introducing source inversion capabilities to SAWLENS2 would vastly increase the codes usefulness in informing statistically more precise parametric methods on such potential systematic errors. A free-form assessment of lens morphology could also be suited to extend the scope of source inversion methods to giant arcs, allowing to study the properties of highly magnified galaxies at extremely high redshifts. But as preliminary attempts have proven, a fully free-form treatment of source and lens faces severe numerical challenges, that may require a more careful treatment than a straightforward translation of currently used techniques to the SAWLENS2 framework. Alternatively, free-form techniques like SAWLENS2 could still be used to inform the choice of parametric model in semi-parametric source inversions.

The adaptable free-form character of the presented framework allows its application also to structures beyond clusters of galaxies. Ongoing work is directed to extending its scope to the cosmic web itself, a transition that has technical but little conceptual challenges in the realm of gravitational lensing but would have to alter the way non-lensing constraints are incorporated.

All in all, the SAWLENS2-framework has been developed and revisited throughout the span of more than two decades, but it seems that there is still room for more.

Appendix A

Additional figures

A.1 Reconstructing a NFW halo

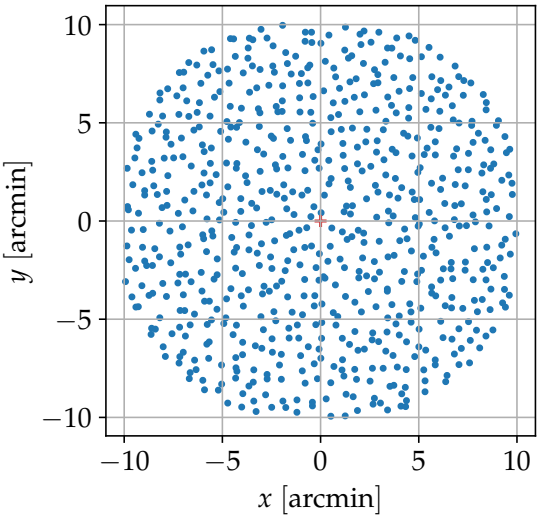

FIGURE A.1: The mesh-free reconstruction domain generated from a shear catalogue of a NFW mock cluster; the red cross marks the center of the coordinate frame and will do so throughout the following maps.

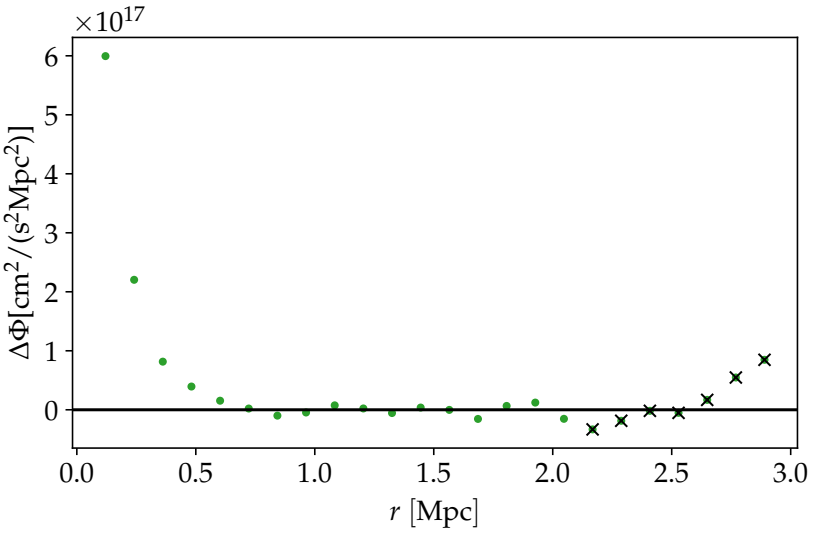

FIGURE A.2: The Laplacian of the Newtonian potential profile; As this is proportional to the total density and therefore has to be non-negative, the crossed out points fail either one or both the sanity checks and are excluded from the analysis.

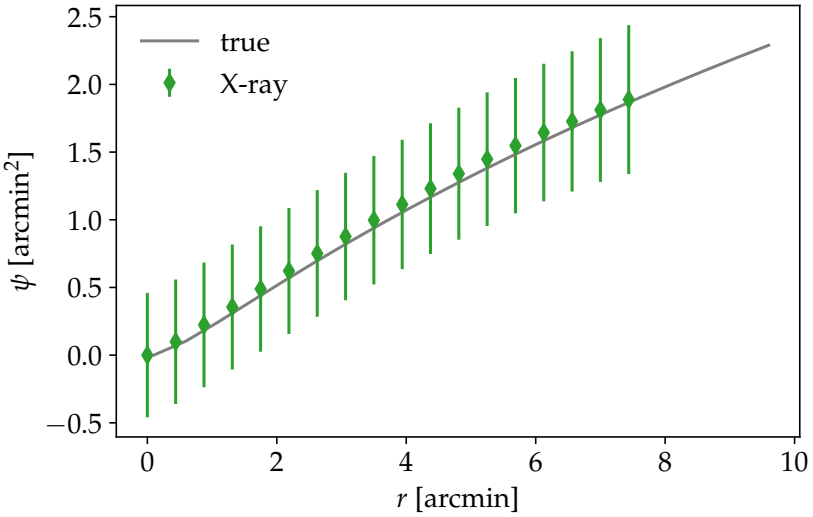

FIGURE A.3: The lensing potential of a NFW mock cluster, obtained from projecting the X-ray based estimate of the Newtonian potential;

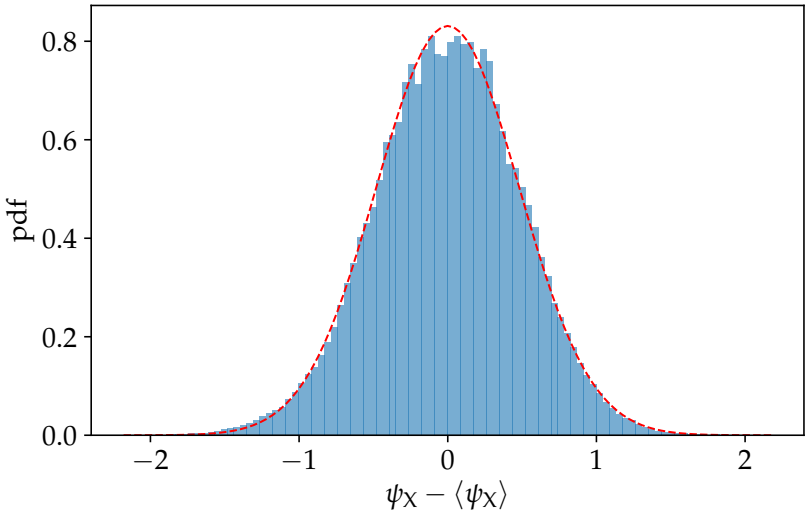

FIGURE A.4: Histogram of the difference between realisations and the original of the lensing potential obtained from projecting the X-ray based estimate of the Newtonian potential of a NFW mock cluster; as the overlayed (red) Gaussian shows, the realisations clearly follow a normal distribution

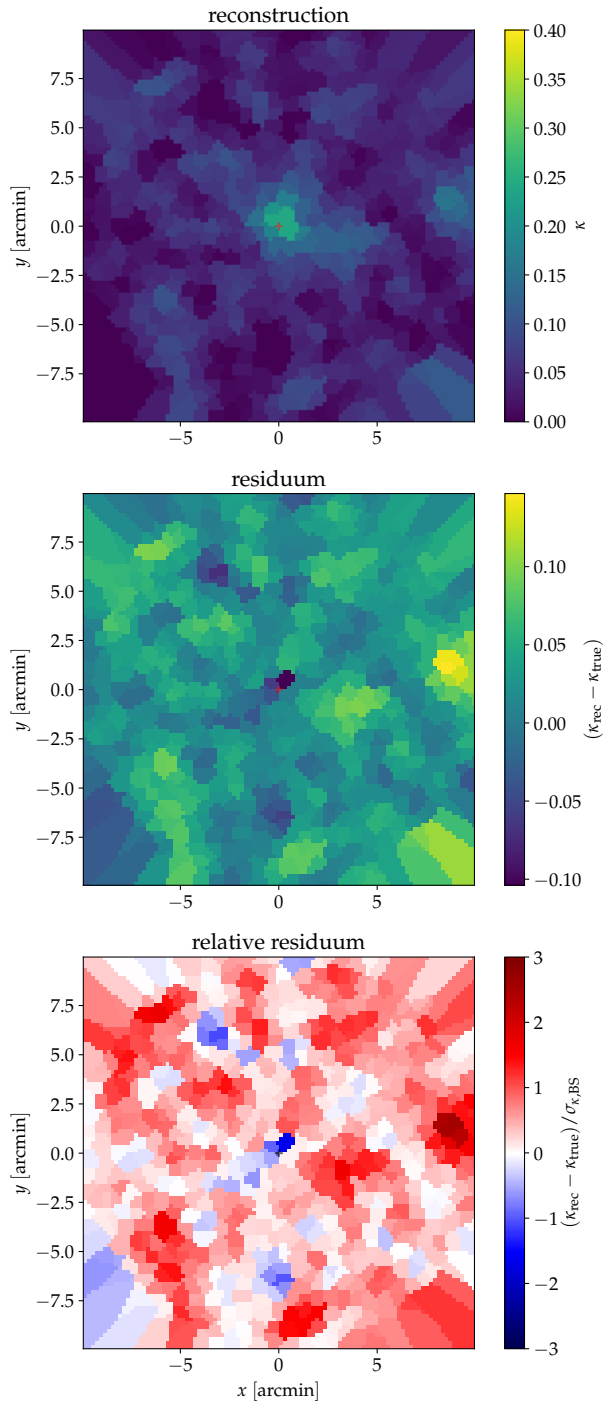

FIGURE A.5: (*top*) Mean reconstructed convergence map for a NFW mock cluster, based on weak lensing alone, inferred from averaging over bootstrapped samples; (*middle*) residuum with respect to true convergence map; (*bottom*) relative residuum, compared to error estimate based on bootstrapping;

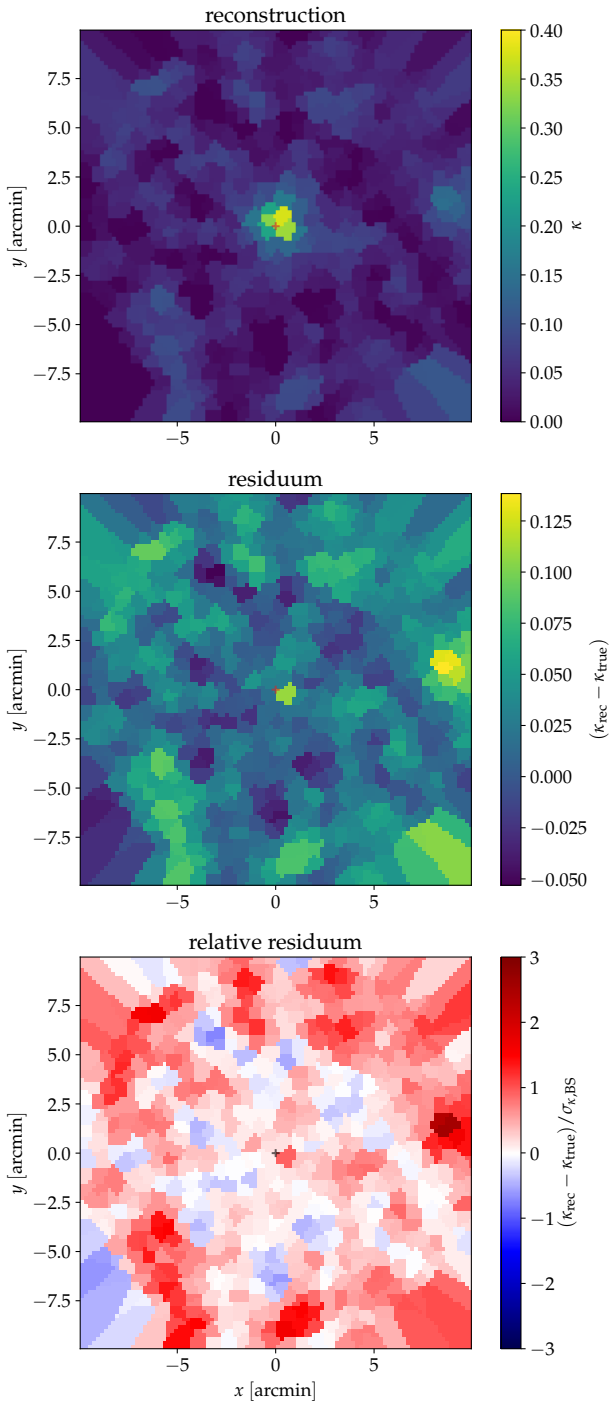

FIGURE A.6: (*top*) Mean reconstructed convergence map for a NFW mock cluster, based on weak lensing and X-ray data, inferred from averaging over bootstrapped samples; (*middle*) residuum with respect to true convergence map; (*bottom*) relative residuum, compared to error estimate based on bootstrapping;

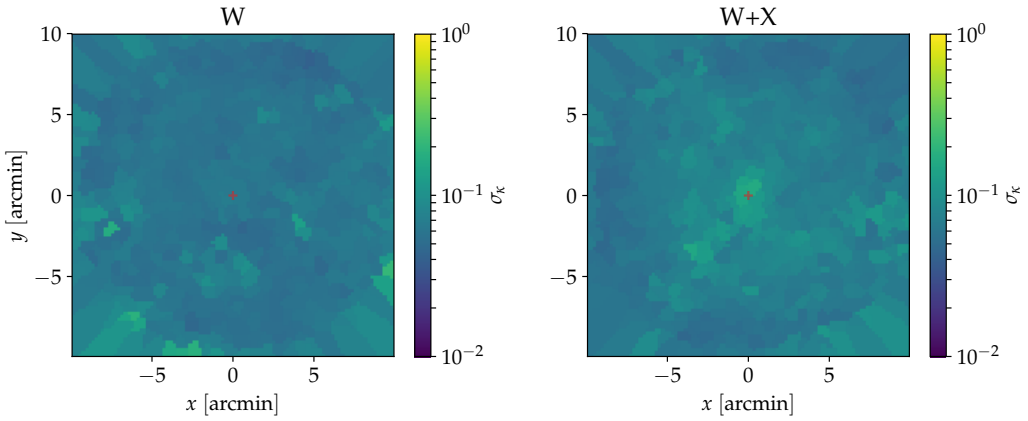

FIGURE A.7: Map of the standard deviation on the reconstructed convergence of a NFW mock cluster, inferred from bootstrapping; (*left*) weak lensing only, (*right*) weak lensing and X-ray data;

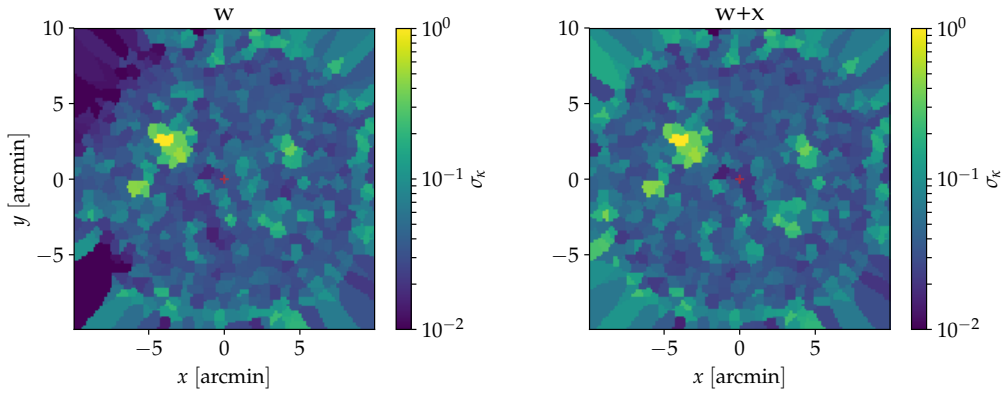

FIGURE A.8: Map of the standard deviation on the reconstructed convergence of a NFW mock cluster, inferred from the Fisher matrix; (*left*) weak lensing only, (*right*) weak lensing and X-ray data;

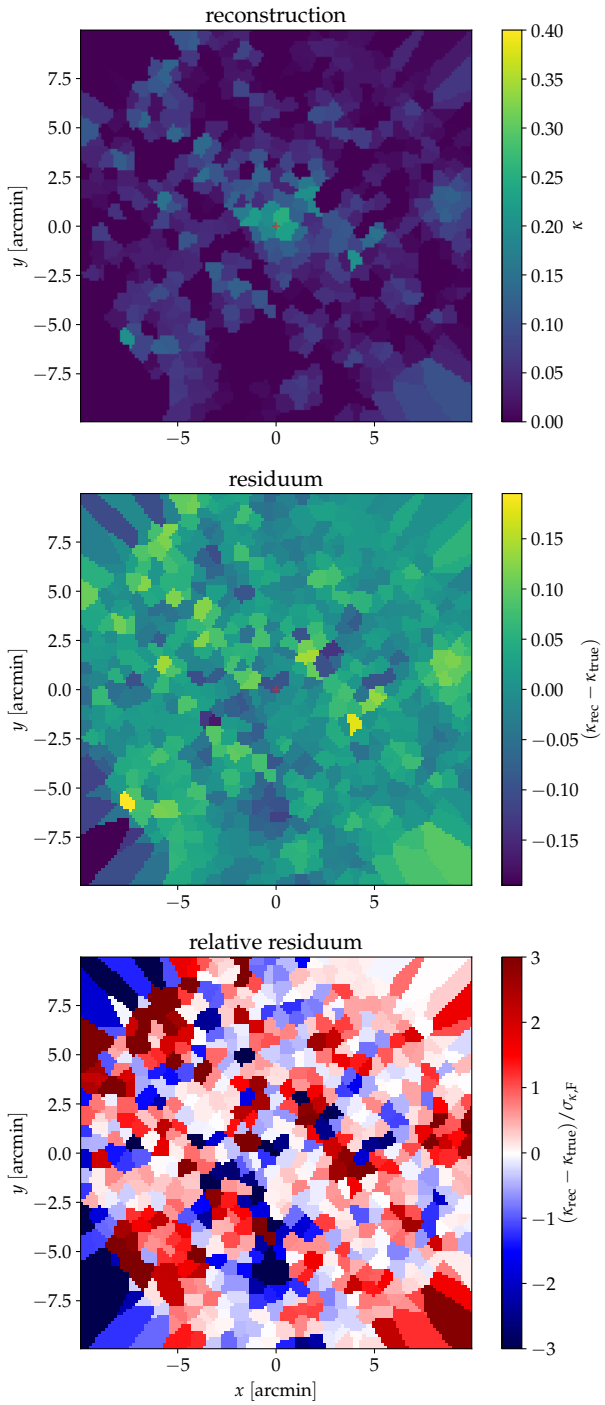

FIGURE A.9: (*top*) Reconstructed convergence map for a NFW mock cluster, based on weak lensing alone; (*middle*) residuum with respect to true convergence map; (*bottom*) relative residuum, compared to error estimate based on the Fisher matrix;

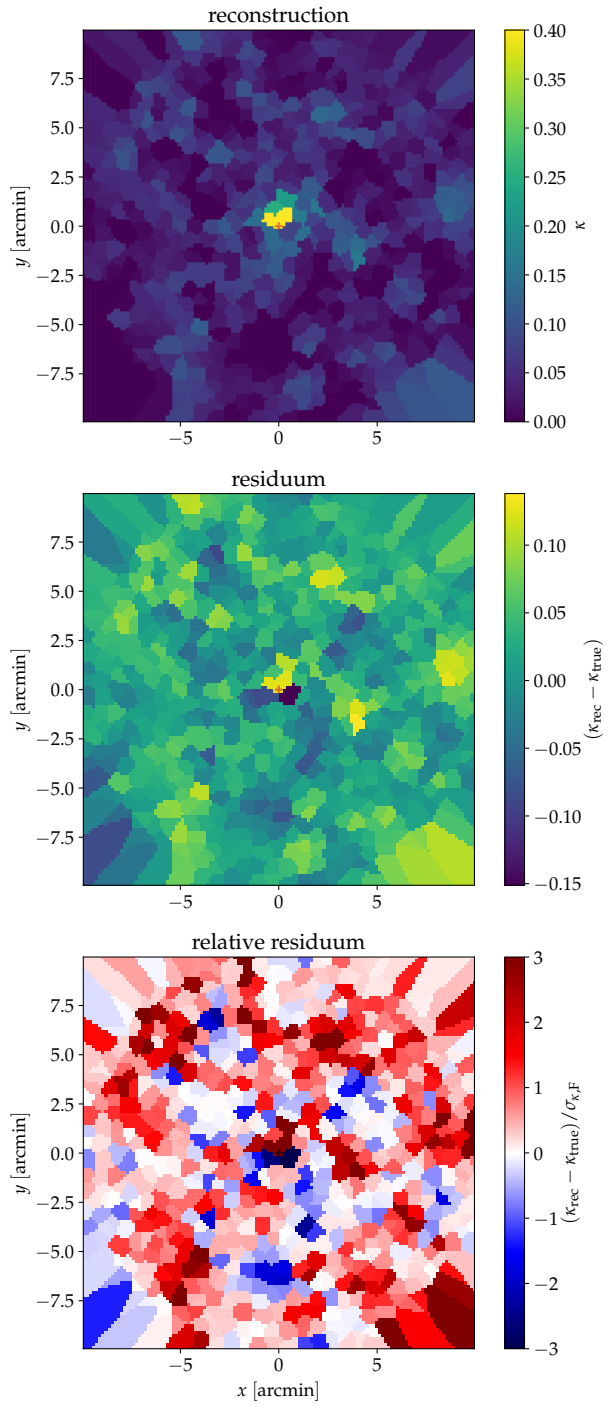

FIGURE A.10: (*top*) Reconstructed convergence map for a NFW mock cluster, based on weak lensing aand X-ray data; (*middle*) residuum with respect to true convergence map; (*bottom*) relative residuum, compared to error estimate based on the Fisher matrix;

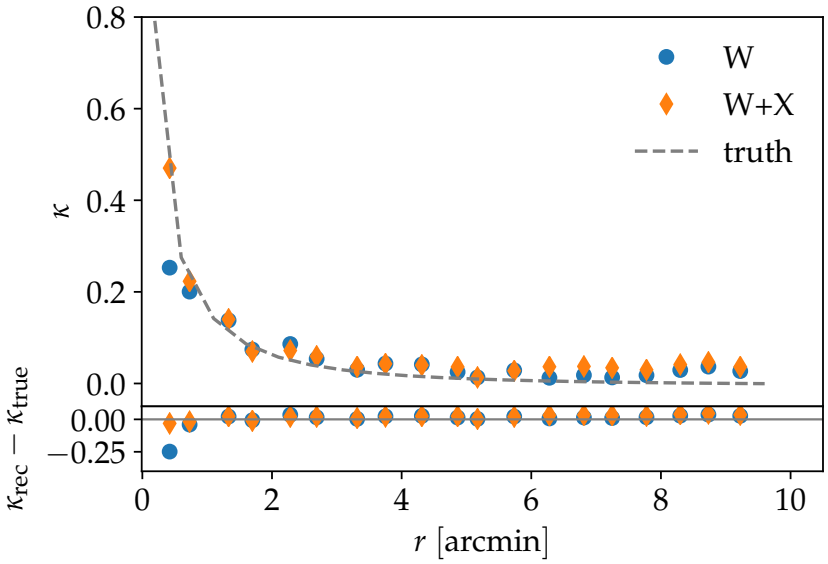

FIGURE A.11: Reconstructed convergence profiles for a NFW mock cluster;

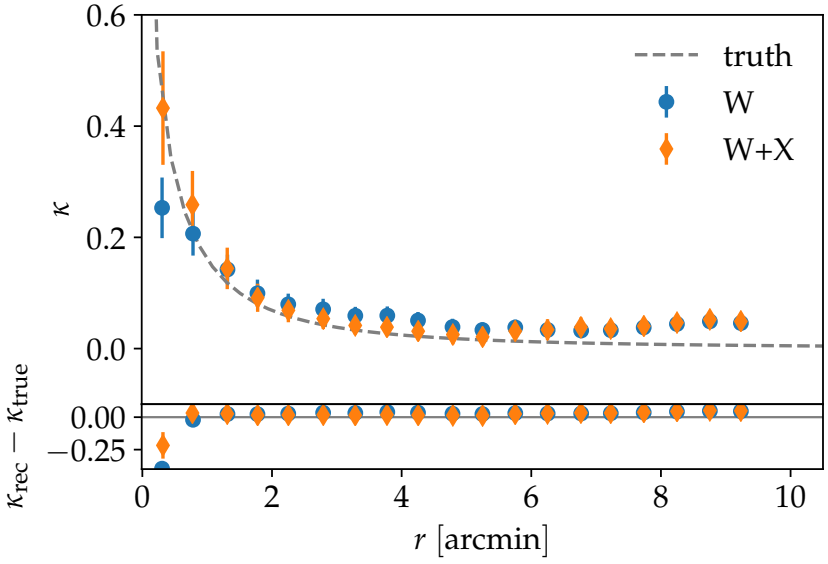

FIGURE A.12: Reconstructed convergence profiles for a NFW mock cluster, inferred from averaging over bootstrapped samples;

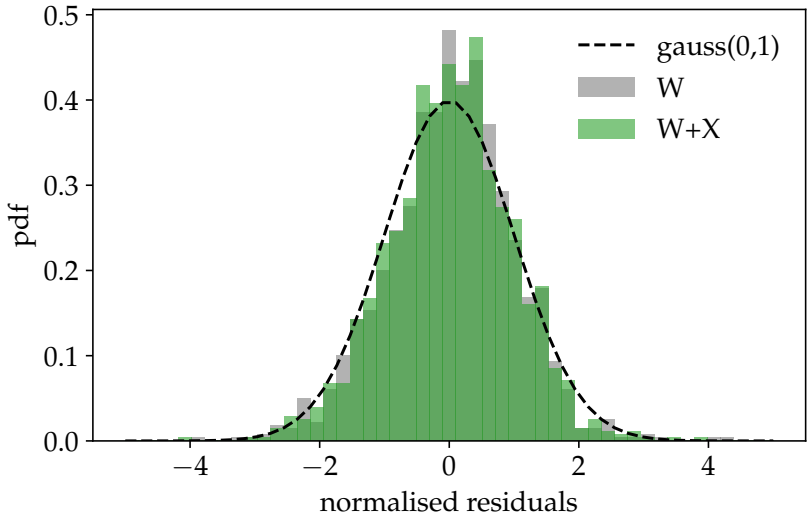

FIGURE A.13: Histogram of the normalized data residuals in the eigenbasis of the inverse data covariance in the case of a NFW halo. The results for the recosntruction using only shear (grey) and using shear and X-ray data (green) are reasonably consistent with a Gaussian of zero mean and unit variance (dashed line).

A.2 Reconstructing a realistic cluster

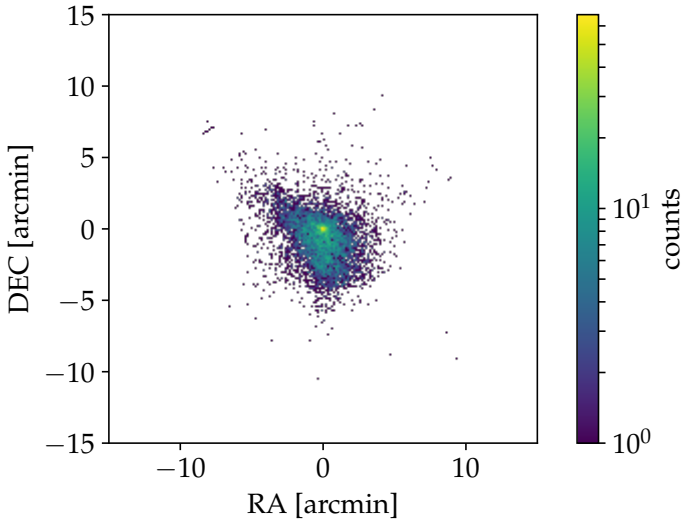

FIGURE A.14: The mock XMM count map for a realistic cluster.

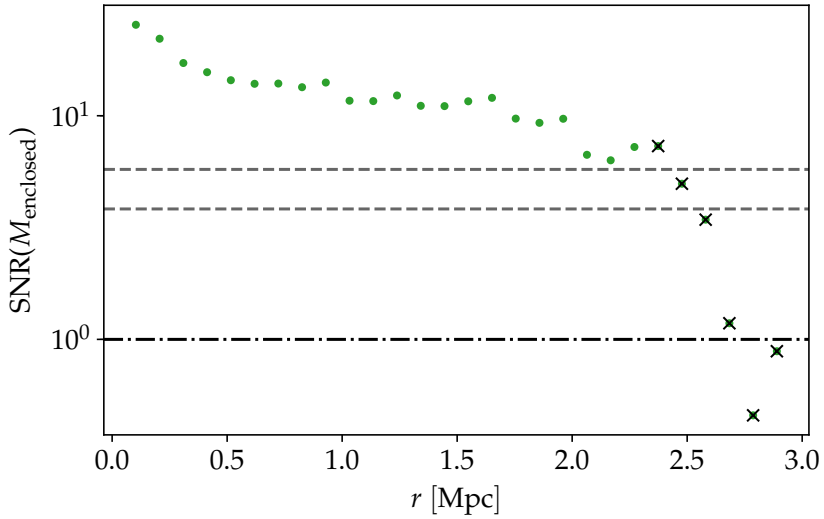

FIGURE A.15: signal-to-noise ratio for the enclosed mass as a function of radial bin for a realistic cluster. Horizontal lines denote signal-to-noise values of 1, 3 and 5 respectively. Crossed-out bins fail either one or both of the sanity checks and are excluded from the analysis.

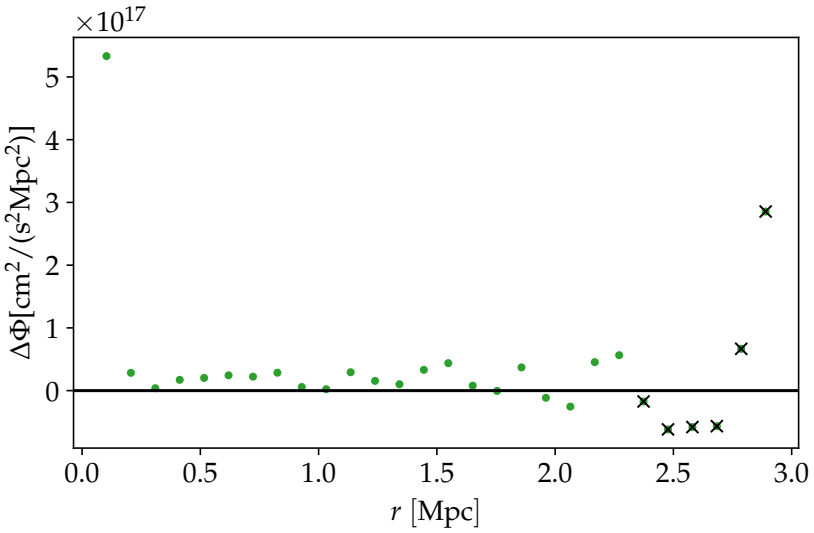

FIGURE A.16: The Laplacian of the Newtonian potential profile for the case of a realistic cluster; As this is proportional to the total density and therefore has to be non-negative, the crossed out points fail either one or both of the sanity checks and are excluded from the analysis.

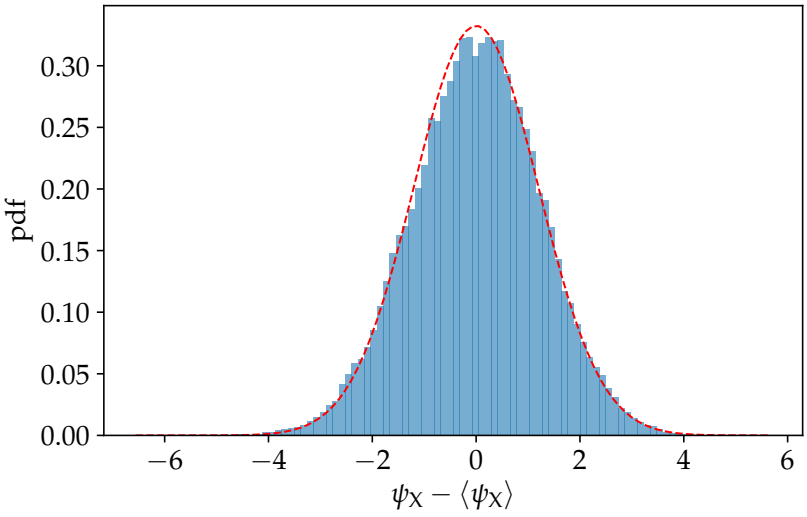

FIGURE A.17: Histogram of the difference between realisations and the original of the lensing potential obtained from projecting the X-ray based estimate of the Newtonian potential of a realistic cluster; as the overlayed (red) Gaussian shows, the realisations clearly follow a normal distribution

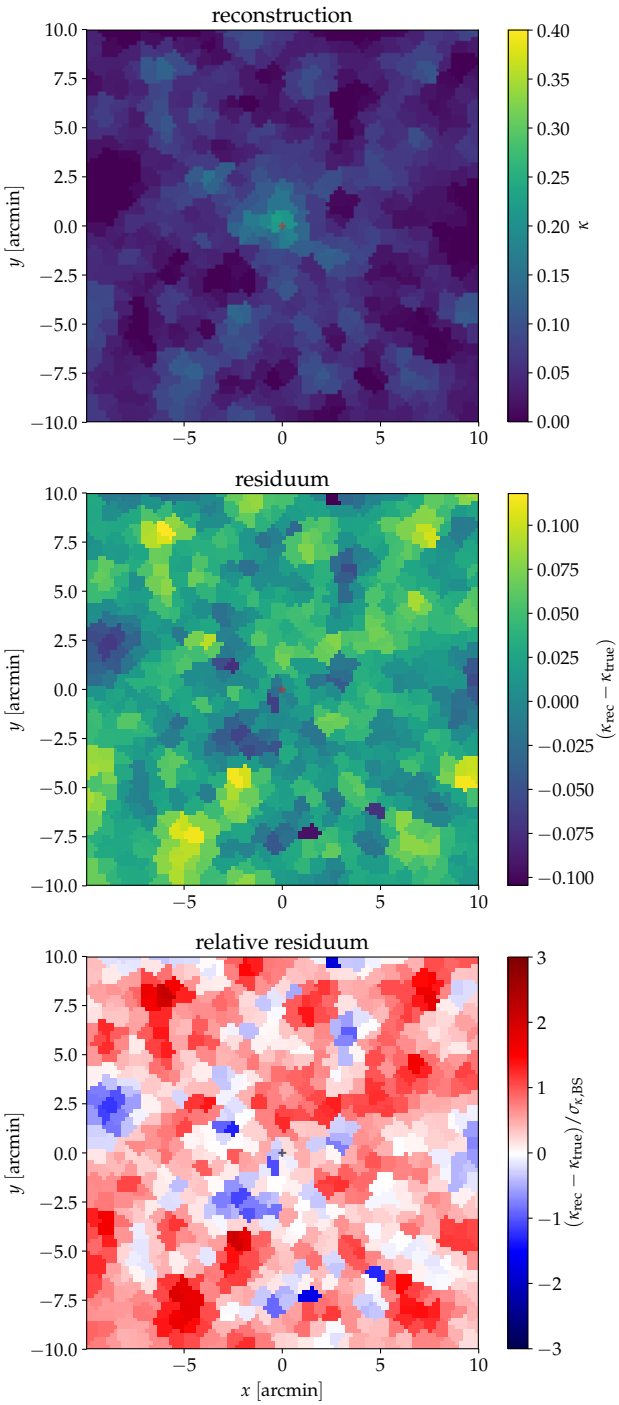

FIGURE A.18: (*top*) Mean reconstructed convergence map for a realistic cluster, based on weak lensing alone, inferred from averaging over bootstrapped samples; (*middle*) residuum with respect to true convergence map; (*bottom*) relative residuum, compared to error estimate based on bootstrapping;

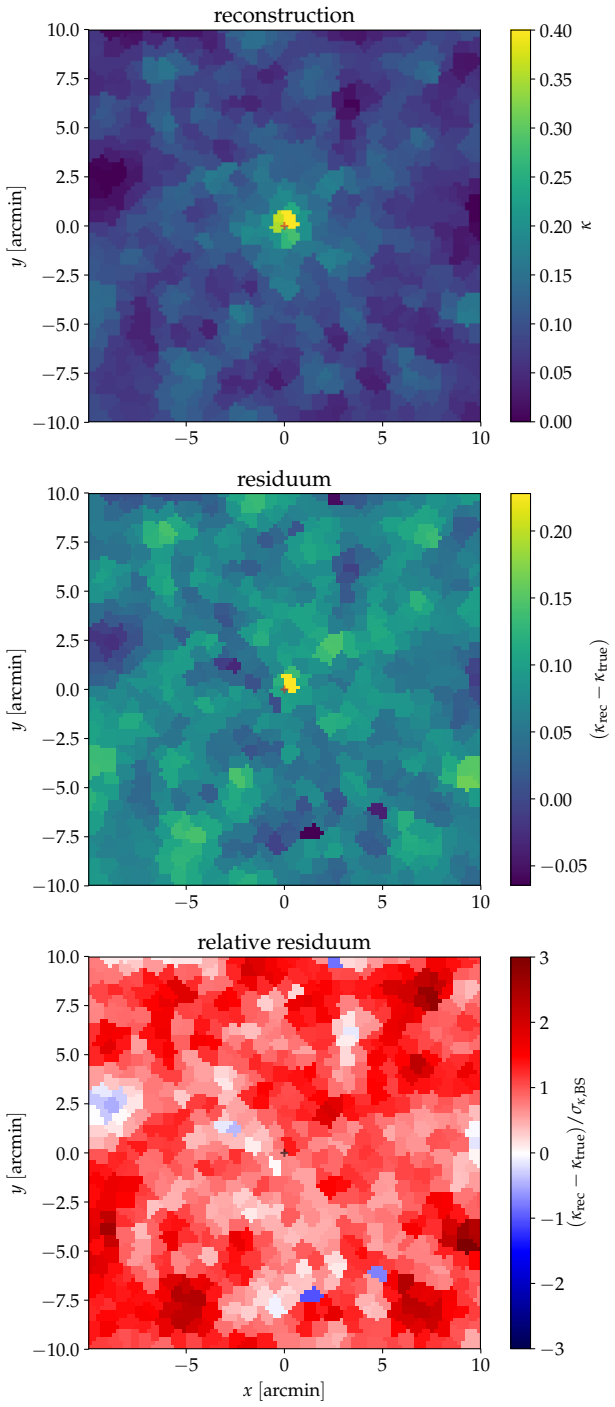

FIGURE A.19: (*top*) Mean reconstructed convergence map for a realistic cluster, based on weak lensing and X-ray data, inferred from averaging over bootstrapped samples; (*middle*) residuum with respect to true convergence map; (*bottom*) relative residuum, compared to error estimate based on bootstrapping;

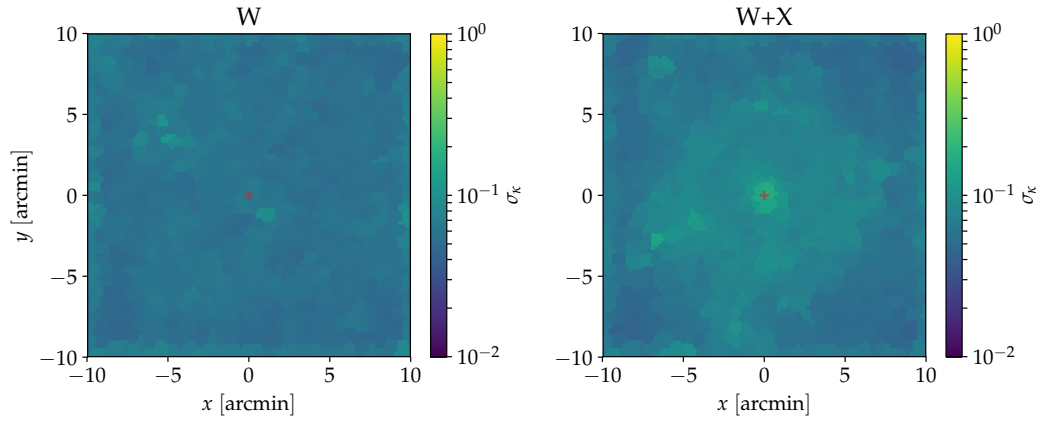

FIGURE A.20: Map of the standard deviation on the reconstructed convergence of a realistic cluster, inferred from bootstrapping; (*left*) weak lensing only, (*right*) weak lensing and X-ray data;

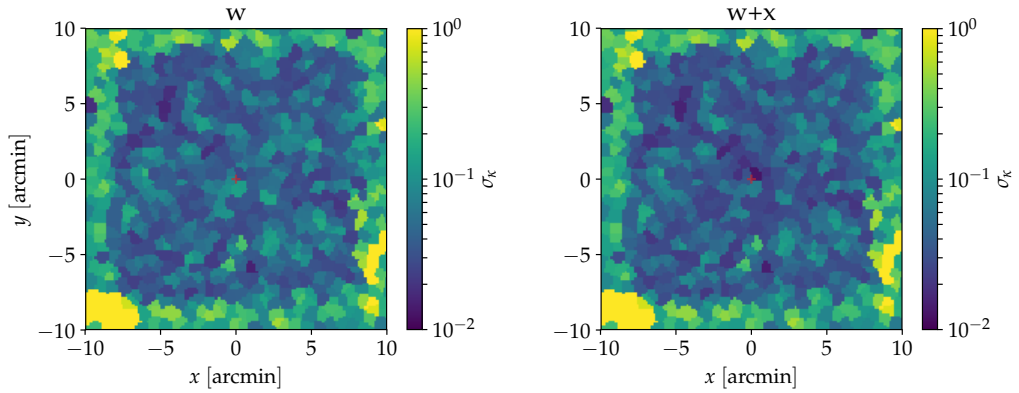

FIGURE A.21: Map of the standard deviation on the reconstructed convergence of a realistic cluster, inferred from the Fisher matrix; (*left*) weak lensing only, (*right*) weak lensing and X-ray data;

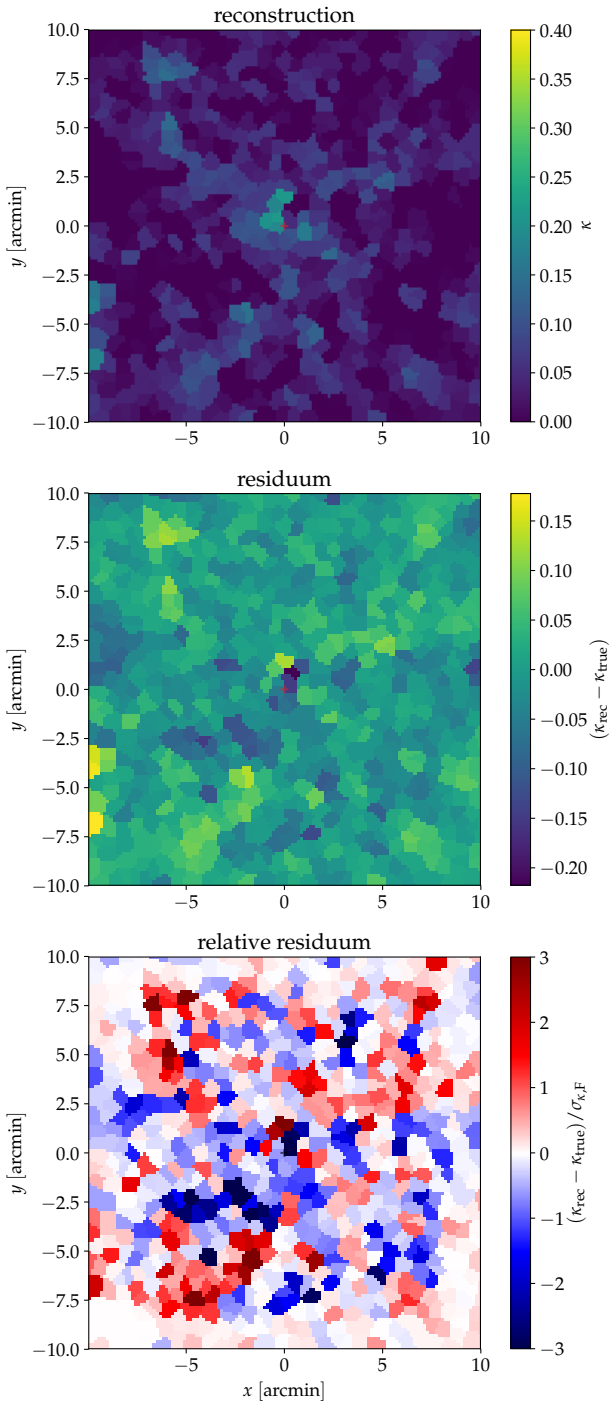

FIGURE A.22: (*top*) Reconstructed convergence map for a realistic cluster, based on weak lensing alone; (*middle*) residuum with respect to true convergence map; (*bottom*) relative residuum, compared to error estimate based on the Fisher matrix;

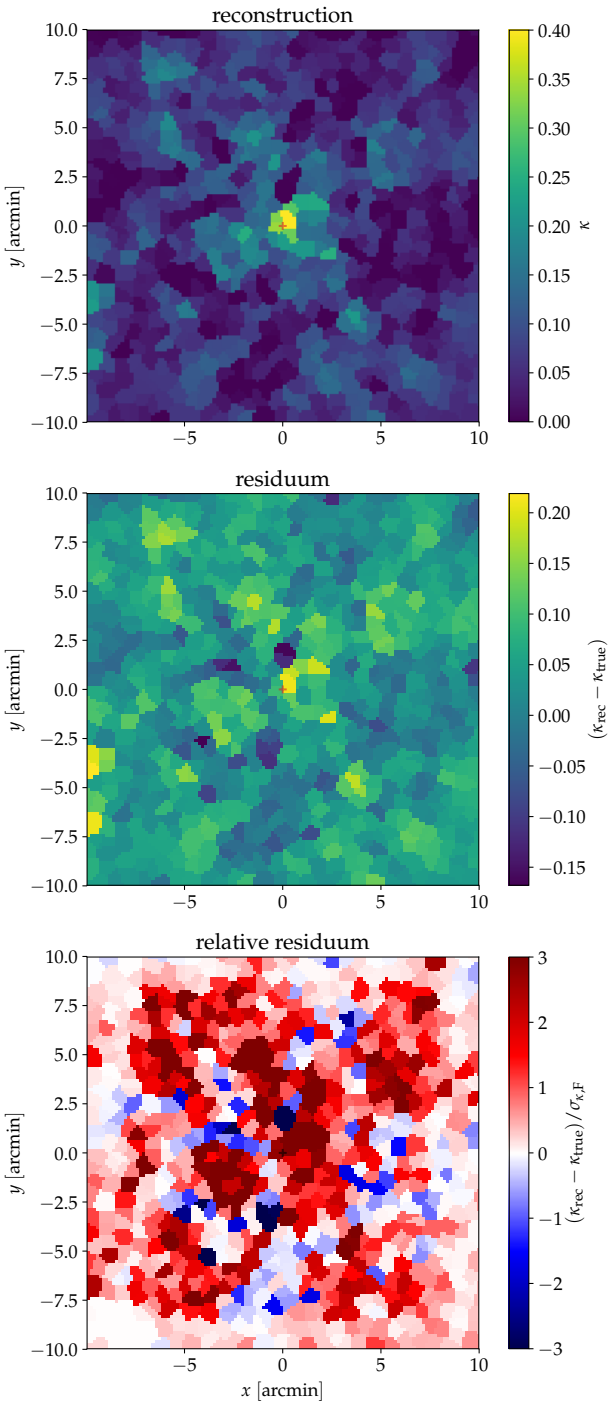

FIGURE A.23: (*top*) Reconstructed convergence map for a realistic cluster, based on weak lensing aand X-ray data; (*middle*) residuum with respect to true convergence map; (*bottom*) relative residuum, compared to error estimate based on the Fisher matrix;

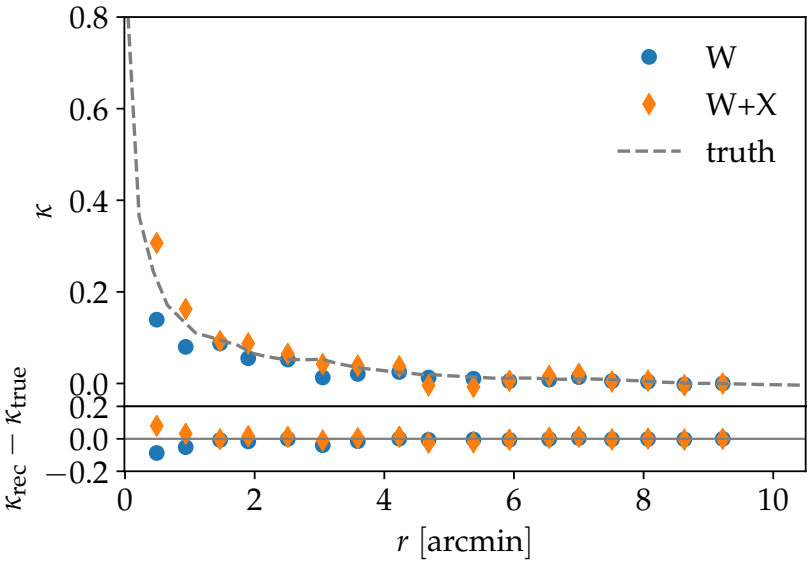

FIGURE A.24: Reconstructed convergence profiles for a realistic cluster; profiles shifted to correct for bias (see Section 4.2);

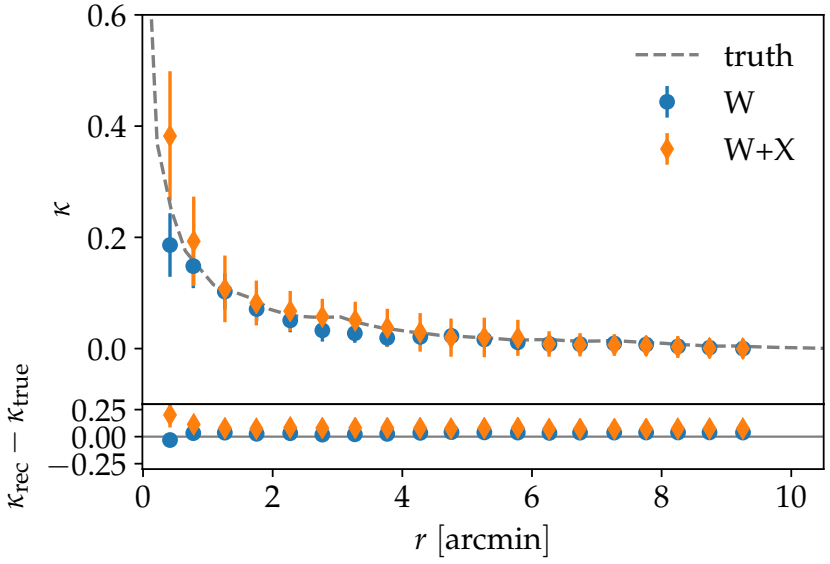

FIGURE A.25: Reconstructed convergence profiles for a realistic cluster, inferred from averaging over bootstrapped samples;profiles shifted to correct for bias (see Section 4.2);

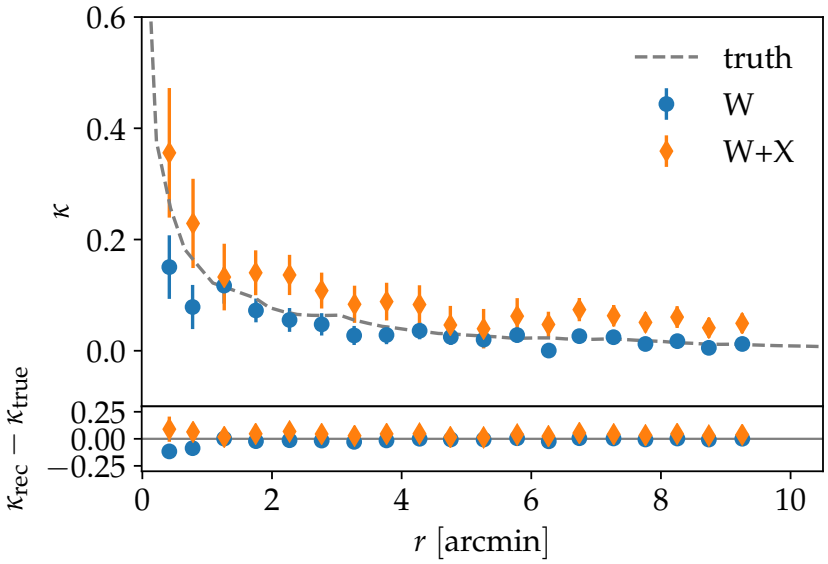

FIGURE A.26: Reconstructed convergence profiles for a realistic cluster, with error bars inferred from averaging over bootstrapped samples; a clear bias towards higher convergence values is visible;

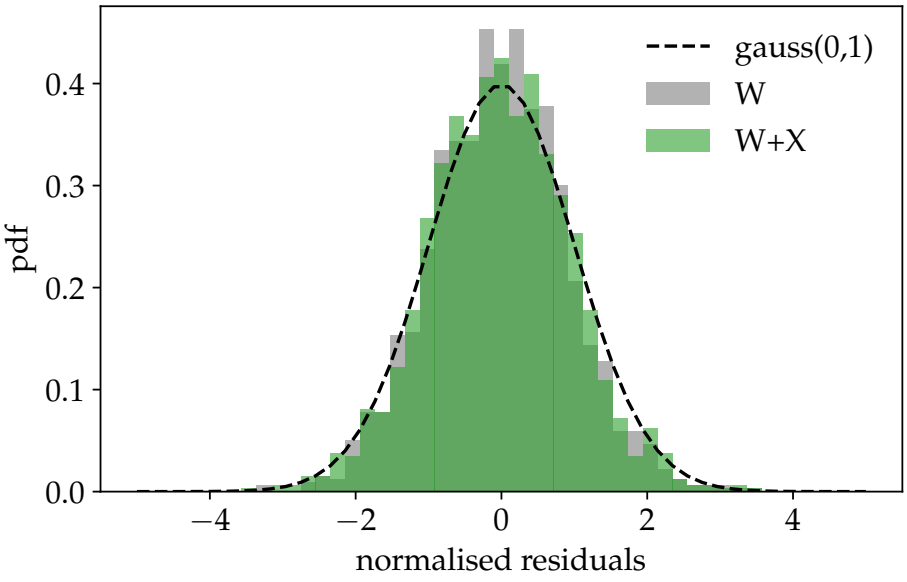

FIGURE A.27: normalized residuals in eigenbasis of the data covariance, for the case of a realistic cluster.The results for the recosntruction using only shear (grey) and using shear and X-ray data (green) are reasonably consistent with a Gaussian of zero mean and unit variance (dashed line).

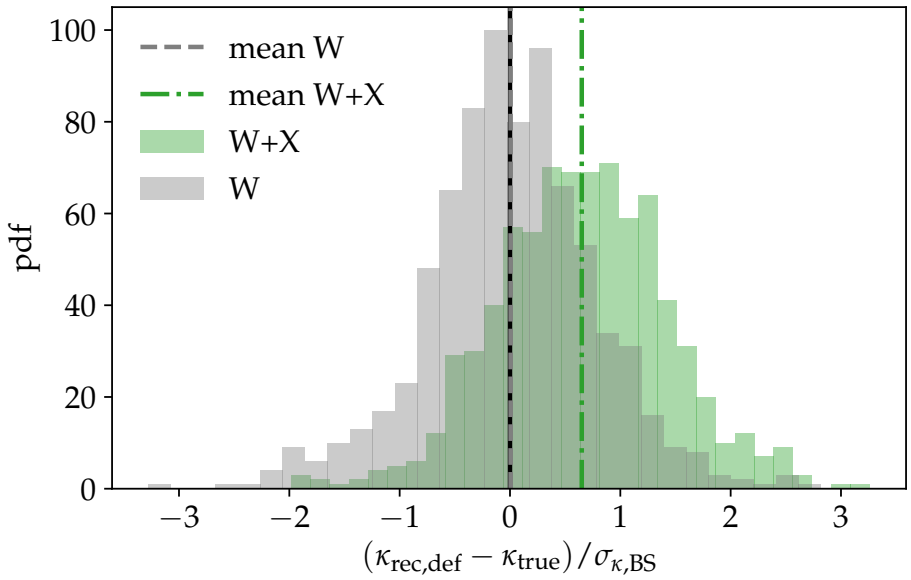

FIGURE A.28: Histogram of the deviation of the default reconstructed convergence map of a realistic cluster from the true map, in units of the bootstrap-based standard deviation. The combined reconstruction is biased to higher convergence values as discussed in section 4.2;

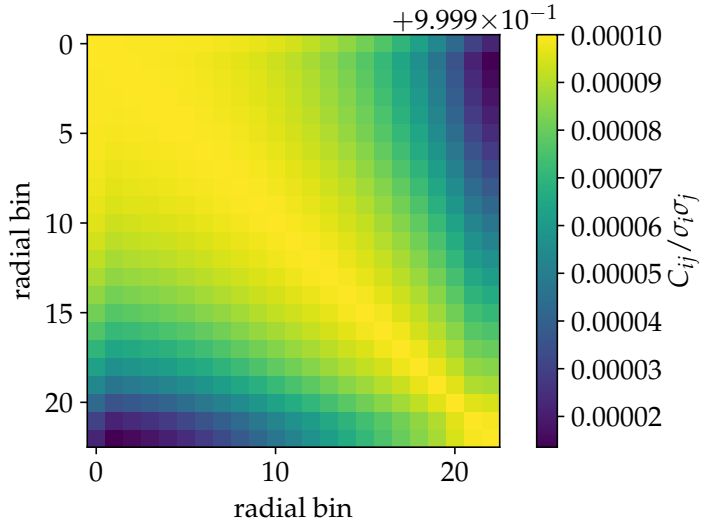

FIGURE A.29: The Pearson correlation matrix of the lensing potential profile realisations recovered from X-ray observations of the realistic cluster; It is clear that, due to the projection step, every bin almost completely correlates with every other;

Appendix B

NFW properties

For the sake of completeness, I give the relevant formulas for the creation of mock observations of the NFW halo in section 4.1. This is based on (Golse and Kneib, 2002) and (Suto, Sasaki, and Makino, 1998; Komatsu and Seljak, 2001). The lensing properties of an NFW halo are given by

$$\psi(x) = 2\kappa_s \theta_s h(x) \tag{B.1}$$
$$\kappa(x) = 2\kappa_s F(x) \tag{B.2}$$
$$\gamma(x) = 2\kappa_s \left(\frac{2}{x^2} G(x) - F(x) \right), \tag{B.3}$$

where we use the dimensionless radius $x = r/r_s \approx \theta/\theta_s$ and

$$\kappa_s = \frac{r_s \rho_s}{\Sigma_{crit}} \tag{B.4}$$
$$F(x) = \begin{cases} \frac{1}{x^2-1}\left(1 - \frac{1}{\sqrt{1-x^2}} \operatorname{arcosh}\frac{1}{x}\right) & \text{if } x < 1 \\ 1/3 & \text{if } x = 1 \\ \frac{1}{x^2-1}\left(1 - \frac{1}{\sqrt{x^2-1}} \arccos\frac{1}{x}\right) & \text{if } x > 1 \end{cases} \tag{B.5}$$
$$G(x) = \begin{cases} \ln\frac{x}{2} + \frac{1}{\sqrt{1-x^2}} \operatorname{arcosh}\frac{1}{x} & \text{if } x < 1 \\ 1 + \ln\frac{1}{2} & \text{if } x = 1 \\ \ln\frac{x}{2} + \frac{1}{\sqrt{x^2-1}} \arccos\frac{1}{x} & \text{if } x > 1 \end{cases} \tag{B.6}$$
$$h(x) = \begin{cases} \left(\ln\frac{x}{2}\right)^2 - \left(\operatorname{arcosh}\frac{1}{x}\right)^2 & \text{if } x < 1 \\ \left(\ln\frac{x}{2}\right)^2 - \left(\arccos\frac{1}{x}\right)^2 & \text{if } x \geq 1 \end{cases}. \tag{B.7}$$

The dimensionless mass profile for the NFW is given by

$$m(x) = \ln(1+x) - \frac{x}{1+x}. \tag{B.8}$$

With this, the hydrostatic equation for the dimensionless gas density profile is solved by

$$\begin{aligned} y_{\mathrm{gas}}(x) &= \left[1 - \frac{3}{\eta_0}\frac{\gamma-1}{\gamma}\frac{c}{m(c)}\int_0^x \mathrm{d}u \frac{m(u)}{u^2}\right]^{\frac{1}{\gamma-1}} \\ &= \left[1 - \frac{3}{\eta_0}\frac{\gamma-1}{\gamma}\frac{c}{m(c)}\left(1 - \frac{\ln(1+x)}{x}\right)\right]^{\frac{1}{\gamma-1}}, \end{aligned} \tag{B.9}$$

where the mass-temperature normalisation factor η_0 is found to be well-fit be

$$\eta_0 = 0.00678\,(c-6.5)^2 + 0.206(c-6.5) + 2.48. \tag{B.10}$$

The surface brightness profile is then proportional to

$$S_X(x) \propto \int_{-\infty}^{\infty} \mathrm{d}l \left[y_{\mathrm{gas}}(\sqrt{x^2+l^2})\right]^{\frac{\gamma+3}{2}} \tag{B.11}$$

and the normalisation factors are found according to Eq. (2.7) and section 3.5.2, with the scale temperature

$$T_0 = \frac{G m_{\mathrm{p}} \eta_0 \mu M_{\mathrm{vir}}}{3 R_{\mathrm{vir}}}, \tag{B.12}$$

where $\mu \approx 0.59$ is the mean molecular weight of the gas.

Bibliography

Abell, George O., Harold G. Corwin, and Ronald P. Olowin (May 1989). "A Catalog of Rich Clusters of Galaxies". en. In: *Astrophys. J. Suppl. Ser.* 70, p. 1. ISSN: 0067-0049. DOI: 10.1086/191333.

Alcock, C. et al. (Oct. 2000). "The MACHO Project: Microlensing Results from 5.7 Years of Large Magellanic Cloud Observations". en. In: *Astrophys. J.* 542.1, p. 281. DOI: 10.1086/309512.

Allen, Steven W., August E. Evrard, and Adam B. Mantz (Sept. 2011). "Cosmological Parameters from Observations of Galaxy Clusters". In: *Annu. Rev. Astron. Astrophys.* 49, pp. 409–470. ISSN: 0066-4146. DOI: 10.1146/annurev-astro-081710-102514.

Amendola, Luca and Shinji Tsujikawa (2010). *Dark Energy: Theory and Observations*.

Andrae, Rene (Sept. 2010). "Error Estimation in Astronomy: A Guide". In: *ArXiv e-prints* 1009, arXiv:1009.2755.

Andrae, Rene, Tim Schulze-Hartung, and Peter Melchior (Dec. 2010). "Dos and Don'ts of Reduced Chi-Squared". In: *ArXiv e-prints* 1012, arXiv:1012.3754.

Bardeen, James M. (Oct. 1980). "Gauge-Invariant Cosmological Perturbations". In: *Physical Review D* 22, pp. 1882–1905. ISSN: 1550-7998. DOI: 10.1103/PhysRevD.22.1882.

Bartelmann, M. and P. Schneider (Jan. 2001). "Weak Gravitational Lensing". In: *Phys. Rep.* 340, pp. 291–472. ISSN: 0370-1573. DOI: 10.1016/S0370-1573(00)00082-X.

Bartelmann, Matthias (Dec. 2010). "TOPICAL REVIEW Gravitational Lensing". In: *Classical and Quantum Gravity* 27, p. 233001. ISSN: 0264-9381. DOI: 10.1088/0264-9381/27/23/233001.

— (2012). *Theoretical Astrophysics: An Introduction*. Wiley-VCH. ISBN: 978-3-527-41004-0.

Bartelmann, Matthias et al. (June 1996). "Maximum-Likelihood Cluster Reconstruction". In: *The Astrophysical Journal Letters* 464, p. L115. ISSN: 0004-637X. DOI: 10.1086/310114.

Bartelmann, Matthias et al. (2016). "A Microscopic, Non-Equilibrium, Statistical Field Theory for Cosmic Structure Formation". en. In: *New J. Phys.* 18.4, p. 043020. ISSN: 1367-2630. DOI: 10.1088/1367-2630/18/4/043020.

Baumann, Daniel (Oct. 2007). "On the Quantum Origin of Structure in the Inflationary Universe". In: *ArXiv e-prints* 0710, arXiv:0710.3187.

Bayes, Thomas (Jan. 1763). "LII. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S". en. In: *Phil. Trans.* 53, pp. 370–418. ISSN: 0261-0523, DOI: 10.1098/rstl.1763.0053.

Beck, A. M. et al. (Jan. 2016). "An Improved SPH Scheme for Cosmological Simulations". In: *Monthly Notices of the Royal Astronomical Society* 455, pp. 2110–2130. ISSN: 0035-8711. DOI: 10.1093/mnras/stv2443.

Bekenstein, Jacob D. (Oct. 2004). "Relativistic Gravitation Theory for the Modified Newtonian Dynamics Paradigm". en. In: *Phys. Rev. D* 70.8, p. 083509. DOI: 10.1103/PhysRevD.70.083509.

Berger, Marsha J. and Joseph Oliger (Mar. 1984). "Adaptive Mesh Refinement for Hyperbolic Partial Differential Equations". en. In: *J. Comput. Phys.* 53.3, p. 484. DOI: 10.1016/0021-9991(84)90073-1.

Bernardeau, F. et al. (Sept. 2002). "Large-Scale Structure of the Universe and Cosmological Perturbation Theory". In: *Physics Reports* 367, pp. 1–248. ISSN: 0370-1573. DOI: 10.1016/S0370-1573(02)00135-7.

Bertone, Gianfranco (2010). "Particle Dark Matter : Observations, Models and Searches". en. In: *Part. Dark Matter Obs. Models Searches*.

Bertschinger, Edmund (1998). "Simulations of Structure Formation in the Universe". en. In: *Annu. Rev. Astron. Astrophys.* 36, p. 599. DOI: 10.1146/annurev.astro.36.1.599.

Biffi, V. et al. (Mar. 2012). "Observing Simulated Galaxy Clusters with PHOX: A Novel X-Ray Photon Simulator". In: *Monthly Notices of the Royal Astronomical Society* 420, pp. 3545–3556. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2011.20278.x.

Biffi, V. et al. (Aug. 2016). "On the Nature of Hydrostatic Equilibrium in Galaxy Clusters". en. In: *Astrophys. J.* 827.2, p. 112. DOI: 10.3847/0004-637X/827/2/112.

Binney, James and Scott Tremaine (1987). *Galactic Dynamics*. Princeton Series in Astrophysics.

Bishop, Christopher (2006). *Pattern Recognition and Machine Learning*. en. Information Science and Statistics. New York: Springer-Verlag. ISBN: 978-0-387-31073-2.

Blandford, Roger and Ramesh Narayan (Nov. 1986). "Fermat's Principle, Caustics, and the Classification of Gravitational Lens Images". en. In: *Astrophys. J.* 310, p. 568. ISSN: 0004-637X. DOI: 10.1086/164709.

Böhringer, Hans and Norbert Werner (Feb. 2010). "X-Ray Spectroscopy of Galaxy Clusters: Studying Astrophysical Processes in the Largest Celestial Laboratories". In: *Astronomy and Astrophysics Review* 18, pp. 127–196. ISSN: 0935-4956. DOI: 10.1007/s00159-009-0023-3.

Bond, J. R. et al. (Oct. 1991). "Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations". en. In: *Astrophys. J.* 379, pp. 440–460. ISSN: 0004-637X. DOI: 10.1086/170520.

Boylan-Kolchin, Michael et al. (Sept. 2009). "Resolving Cosmic Structure Formation with the Millennium-II Simulation". In: *Mon. Not. R. Astron. Soc.* 398, pp. 1150–1164. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2009.15191.x.

Bradač, M. et al. (July 2005). "Strong and Weak Lensing United. I. The Combined Strong and Weak Lensing Cluster Mass Reconstruction Method". en. In: *Astron. Astrophys.* 437.1, p. 39. DOI: 10.1051/0004-6361:20042233.

Buchmueller, O. et al. (June 2014). "The CMSSM and NUHM1 after LHC Run 1". In: *European Physical Journal C* 74, p. 2922. ISSN: 1434-6044. DOI: 10.1140/epjc/s10052-014-2922-3.

Burke, W. L. (Feb. 1981). "Multiple Gravitational Imaging by Distributed Masses". en. In: *Astrophys. J.* 244, p. L1. DOI: 10.1086/183466.

Cacciato, M. et al. (Nov. 2006). "Combining Weak and Strong Lensing in Cluster Potential Reconstruction". In: *Astron. Astrophys.* 458, pp. 349–356. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20054582.

Capasso, R. et al. (Oct. 2018). "Galaxy Kinematics and Mass Calibration in Massive SZE Selected Galaxy Clusters to Z=1.3". en. In: *Mon. Not. R. Astron. Soc.* P. 2531. DOI: 10.1093/mnras/sty2645.

Carlstrom, John E., Gilbert P. Holder, and Erik D. Reese (2002). "Cosmology with the Sunyaev-Zel'dovich Effect". In: *Annual Review of Astronomy and Astrophysics* 40, pp. 643–680. ISSN: 0066-4146. DOI: 10.1146/annurev.astro.40.060401.093803.

Clowe, Douglas et al. (Sept. 2006). "A Direct Empirical Proof of the Existence of Dark Matter". In: *Astrophys. J. Lett.* 648, pp. L109–L113. ISSN: 0004-637X. DOI: 10.1086/508162.

Coe, Dan (May 2010). "Dark Matter Halo Mass Profiles". In: *ArXiv e-prints* 1005, arXiv:1005.0411.

Daubechies, Ingrid (Sept. 1990). "The Wavelet Transform, Time-Frequency Localization and Signal Analysis". en. In: *IEEE Trans. Inf. Theory* 36, p. 961. ISSN: 0018-9448.

Davis, M. et al. (May 1985). "The Evolution of Large-Scale Structure in a Universe Dominated by Cold Dark Matter". en. In: *Astrophys. J.* 292, p. 371. ISSN: 0004-637X. DOI: 10.1086/163168.

Despali, Giulia et al. (Apr. 2018). "Modelling the Line-of-Sight Contribution in Substructure Lensing". In: *Monthly Notices of the Royal Astronomical Society* 475, pp. 5424–5442. ISSN: 0035-8711. DOI: 10.1093/mnras/sty159.

Dolag, K. et al. (Oct. 2005). "The Imprints of Local Superclusters on the Sunyaev-Zel'dovich Signals and Their Detectability with Planck". In: *Monthly Notices of the Royal Astronomical Society* 363, pp. 29–39. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09452.x.

Dolag, K. et al. (Oct. 2009). "Substructures in Hydrodynamical Cluster Simulations". en. In: *Mon. Not. R. Astron. Soc.* 399.2, p. 497. DOI: 10.1111/j.1365-2966.2009.15034.x.

Doroshkevich, A. G. (Oct. 1970). "Spatial Structure of Perturbations and Origin of Galactic Rotation in Fluctuation Theory". en. In: *Astrophysics* 6.4, p. 320. DOI: 10.1007/BF01001625.

Duffy, Alan R. et al. (Oct. 2008). "Dark Matter Halo Concentrations in the Wilkinson Microwave Anisotropy Probe Year 5 Cosmology". en. In: *Mon. Not. R. Astron. Soc.* 390.1, p. L64. DOI: 10.1111/j.1745-3933.2008.00537.x.

Einasto, J. and U. Haud (Oct. 1989). "Galactic Models with Massive Corona. I - Method. II - Galaxy". In: *Astron Astrophys* 223, pp. 89–106.

Einstein, A. (1916). "Die Grundlage Der Allgemeinen Relativitätstheorie". en. In: *Ann. Phys.* 354.7, p. 769. DOI: 10.1002/andp.19163540702.

Einstein, Albert (1915). "Zur Allgemeinen Relativitätstheorie". en. In: *Sitzungsberichte K. Preußischen Akad. Wiss. Berl.* P. 778.

Evrard, A. E. and J. P. Henry (Dec. 1991). "Expectations for X-Ray Cluster Observations by the ROSAT Satellite". en. In: *Astrophys. J.* 383, p. 95. ISSN: 0004-637X. DOI: 10.1086/170767.

Farrar, Glennys R. and P. J. E. Peebles (Mar. 2004). "Interacting Dark Matter and Dark Energy". en. In: *Astrophys. J.* 604.1, p. 1. DOI: 10.1086/381728.

Flyer, Natasha et al. (Sept. 2016). "On the Role of Polynomials in RBF-FD Approximations: I. Interpolation and Accuracy". In: *J. Comput. Phys.* 321, p. 21. DOI: 10.1016/j.jcp.2016.05.026.

Forman, W. et al. (Dec. 2005). "Reflections of Active Galactic Nucleus Outbursts in the Gaseous Atmosphere of M87". en. In: *Astrophys. J.* 635.2, p. 894. DOI: 10.1086/429746.

Fornberg, Bengt (Jan. 1998). "Classroom Note:Calculation of Weights in Finite Difference Formulas". en. In: *SIAM Rev.* 40.3, p. 685. DOI: 10.1137/S0036144596322507.

Fornberg, Bengt, Erik Lehto, and Collin Powell (Feb. 2013). "Stable Calculation of Gaussian-Based RBF-FD Stencils". In: *Computers & Mathematics with*

Applications 65.4, pp. 627–637. ISSN: 0898-1221. DOI: 10.1016/j.camwa.2012.11.006.

Frenk, C. S. and S. D. M. White (Oct. 2012). "Dark Matter and Cosmic Structure". In: *Ann. Phys.* 524, pp. 507–534. ISSN: 0003-3804. DOI: 10.1002/andp.201200212.

Friedmann, A. (1922). "Über Die Krümmung Des Raumes". en. In: *Z. Phys.* 10, p. 377. DOI: 10.1007/BF01332580.

Gingold, R. A. and J. J. Monaghan (Nov. 1977). "Smoothed Particle Hydrodynamics: Theory and Application to Non-Spherical Stars." en. In: *Mon. Not. R. Astron. Soc.* 181, p. 375. DOI: 10.1093/mnras/181.3.375.

Giocoli, Carlo et al. (May 2010). "The Substructure Hierarchy in Dark Matter Haloes". In: *Monthly Notices of the Royal Astronomical Society* 404, pp. 502–517. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2010.16311.x.

Giocoli, Carlo et al. (May 2014). "Mass and Concentration Estimates from Weak and Strong Gravitational Lensing: A Systematic Study". en. In: *Mon. Not. R. Astron. Soc.* 440.2, p. 1899. DOI: 10.1093/mnras/stu303.

Golse, G. and J.-P. Kneib (Aug. 2002). "Pseudo Elliptical Lensing Mass Model: Application to the NFW Mass Distribution". In: *Astronomy and Astrophysics* 390, pp. 821–827. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20020639.

Gorenstein, M. V., E. E. Falco, and I. I. Shapiro (Apr. 1988). "Degeneracies in Parameter Estimates for Models of Gravitational Lens Systems". In: *Astron J* 327, pp. 693–711.

Goto, Tomotsugu et al. (Dec. 2003). "The Morphology-Density Relation in the Sloan Digital Sky Survey". en. In: *Mon. Not. R. Astron. Soc.* 346.2, p. 601. DOI: 10.1046/j.1365-2966.2003.07114.x.

Hagstotz, Steffen et al. (June 2018). "Joint Halo Mass Function for Modified Gravity and Massive Neutrinos I: Simulations and Cosmological Forecasts". In: *ArXiv e-prints* 1806, arXiv:1806.07400.

Hastings, W. K. (Apr. 1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". en. In: *Biometrika* 57.1, pp. 97–109. ISSN: 0006-3444. DOI: 10.1093/biomet/57.1.97.

Hirschmann, Michaela et al. (Aug. 2014). "Cosmological Simulations of Black Hole Growth: AGN Luminosities and Downsizing". In: *Monthly Notices of the Royal Astronomical Society* 442, pp. 2304–2324. ISSN: 0035-8711. DOI: 10.1093/mnras/stu1023.

Hu, Wayne and Scott Dodelson (2002). "Cosmic Microwave Background Anisotropies". In: *Annual Review of Astronomy and Astrophysics* 40, pp. 171–216. ISSN: 0066-4146. DOI: 10.1146/annurev.astro.40.060401.093926.

Huber, Korbinian et al. (Dec. 2018). "Joint Cluster Reconstructions: Combining Free-Form Lensing and X-Rays". en. In: *ArXiv E-Prints*, arXiv:1812.08152.

Hurier, G. and R. E. Angulo (Feb. 2018). "Measuring the Hydrostatic Mass Bias in Galaxy Clusters by Combining Sunyaev-Zel'dovich and CMB Lensing Data". en. In: *Astron. Astrophys.* 610, p. L4. DOI: 10.1051/0004-6361/201731999.

Hütsi, G. and O. Lahav (Dec. 2008). "The Cluster-Galaxy Cross Spectrum. An Additional Probe of Cosmological and Halo Parameters". en. In: *Astron. Astrophys.* 492.2, p. 355. DOI: 10.1051/0004-6361:200810250.

Jauzac, M. et al. (June 2016). "The Extraordinary Amount of Substructure in the Hubble Frontier Fields Cluster Abell 2744". In: *ArXiv e-prints* 1606, arXiv:1606.04527.

Jenkins, A. et al. (Feb. 2001). "The Mass Function of Dark Matter Haloes". In: *Monthly Notices of the Royal Astronomical Society* 321, pp. 372–384. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2001.04029.x.

Kelly, Patrick L. et al. (Apr. 2018). "Extreme Magnification of an Individual Star at Redshift 1.5 by a Galaxy-Cluster Lens". en. In: *Nat. Astron.* 2, p. 334. DOI: 10.1038/s41550-018-0430-3.

Klypin, Anatoly et al. (June 2001). "Resolving the Structure of Cold Dark Matter Halos". In: *The Astrophysical Journal* 554, pp. 903–915. ISSN: 0004-637X. DOI: 10.1086/321400.

Kneib, Jean-Paul et al. (June 2004). "A Probable Z~7 Galaxy Strongly Lensed by the Rich Cluster A2218: Exploring the Dark Ages". en. In: *Astrophys. J.* 607.2, p. 697. DOI: 10.1086/386281.

Komatsu, E. and U. Seljak (Nov. 2001). "Universal Gas Density and Temperature Profile". In: *Monthly Notices of the Royal Astronomical Society* 327, pp. 1353–1366. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2001.04838.x.

Komatsu, E. et al. (Feb. 2011). "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation". In: *The Astrophysical Journal Supplement Series* 192, p. 18. ISSN: 0067-0049. DOI: 10.1088/0067-0049/192/2/18.

Konrad, S. et al. (May 2013). "Joint Reconstruction of Galaxy Clusters from Gravitational Lensing and Thermal Gas. I. Outline of a Non-Parametric Method". In: *Astron. Astrophys.* 553, A118. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201220663.

Koopmans, L. V. E. (Nov. 2005). "Gravitational Imaging of Cold Dark Matter Substructures". In: *Mon. Not. R. Astron. Soc.* 363, pp. 1136–1144. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09523.x.

Kravtsov, Andrey V. and Stefano Borgani (Sept. 2012). "Formation of Galaxy Clusters". In: *Annual Review of Astronomy and Astrophysics* 50, pp. 353–409. ISSN: 0066-4146. DOI: 10.1146/annurev-astro-081811-125502.

Lemaître, G. (1927). "Un Univers Homogène de Masse Constante et de Rayon Croissant Rendant Compte de La Vitesse Radiale Des Nébuleuses Extra-Galactiques". en. In: *Ann. Socieacuteteacute Sci. Brux.* 47, p. 49.

Lesgourgues, Julien and Sergio Pastor (July 2006). "Massive Neutrinos and Cosmology". In: *Physics Reports* 429.6, pp. 307–379. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2006.04.001.

Limousin, Marceau et al. (Aug. 2013). "The Three-Dimensional Shapes of Galaxy Clusters". In: *Space Science Reviews* 177, pp. 155–194. ISSN: 0038-6308. DOI: 10.1007/s11214-013-9980-y.

Liu, G. R. (July 2002). *Mesh Free Methods: Moving Beyond the Finite Element Method*. en. CRC Press. ISBN: 978-1-4200-4058-6.

Liu, G. R., M. B. Liu, and Shaofan Li (2004). ""Smoothed Particle Hydrodynamics - a Meshfree Method"". en. In: *Comput. Mech.* 33.6, p. 491. DOI: 10.1007/s00466-004-0573-1.

Lovelock, David (June 1972). "The Four-Dimensionality of Space and the Einstein Tensor". In: *Journal of Mathematical Physics* 13, pp. 874–876. ISSN: 1527-2427. DOI: 10.1063/1.1666069.

Lucy, L. B. (June 1974). "An Iterative Technique for the Rectification of Observed Distributions". In: *The Astronomical Journal* 79, p. 745. ISSN: 0004-6256. DOI: 10.1086/111605.

— (Dec. 1977). "A Numerical Approach to the Testing of the Fission Hypothesis." en. In: *Astron. J.* 82, p. 1013. DOI: 10.1086/112164.

— (1994). "Optimum Strategies for Inverse Problems in Statistical Astronomy". In: *Astron Astrophys* 289, pp. 983–994.

Macciò, Andrea V. et al. (Dec. 2008). "Concentration, Spin and Shape of Dark Matter Haloes as a Function of the Cosmological Model: WMAP1, WMAP3 and WMAP5 Results". en. In: *Mon. Not. R. Astron. Soc.* 391.4, p. 1940. DOI: 10.1111/j.1365-2966.2008.14029.x.

Majer, C. L. et al. (July 2016). "Reconstruction of the Mass Distribution of Galaxy Clusters from the Inversion of the Thermal Sunyaev-Zel'dovich Effect". In: *Monthly Notices of the Royal Astronomical Society* 460, pp. 844–854. ISSN: 0035-8711. DOI: 10.1093/mnras/stw952.

Majer, Charles Ludwig (2014). *Joint Reconstruction of the Mass Distributions of Galaxy Clusters from Gravitational Lensing and Thermal Gas*. eng. http://archiv.ub.uni-heidelberg.de/volltextserver/17156/. Dissertation. DOI: DOI:10.11588/heidok.00017156.

Mana, Annalisa et al. (Sept. 2013). "Combining Clustering and Abundances of Galaxy Clusters to Test Cosmology and Primordial Non-Gaussianity". en. In: *Mon. Not. R. Astron. Soc.* 434.1, p. 684. DOI: 10.1093/mnras/stt1062.

Markevitch, Maxim (Sept. 1998). "The L_X-T Relation and Temperature Function for Nearby Clusters Revisited". en. In: *Astrophys. J.* 504.1, p. 27. DOI: 10.1086/306080.

Marrodán Undagoitia, Teresa and Ludwig Rauch (Jan. 2016). "Dark Matter Direct-Detection Experiments". In: *Journal of Physics G Nuclear Physics* 43, p. 013001. ISSN: 0954-3899. DOI: 10.1088/0954-3899/43/1/013001.

Mavromatos, Nick E., Mairi Sakellariadou, and Muhammad Furqaan Yusaf (Apr. 2009). "Can the Relativistic Field Theory Version of Modified Newtonian Dynamics Avoid Dark Matter on Galactic Scales?" en. In: *Phys. Rev. D* 79.8, p. 081301. DOI: 10.1103/PhysRevD.79.081301.

McNamara, B. R. and P. E. J. Nulsen (Sept. 2007). "Heating Hot Atmospheres with Active Galactic Nuclei". en. In: *Annu. Rev. Astron. Astrophys.* 45.1, p. 117. DOI: 10.1146/annurev.astro.45.051806.110625.

Meneghetti, M. et al. (Dec. 2014). "The MUSIC of CLASH: Predictions on the Concentration-Mass Relation". In: *The Astrophysical Journal* 797, p. 34. ISSN: 0004-637X. DOI: 10.1088/0004-637X/797/1/34.

Merritt, David et al. (Dec. 2006). "Empirical Models for Dark Matter Halos. I. Nonparametric Construction of Density Profiles and Comparison with Parametric Models". In: *Astron. J.* 132, pp. 2685–2700. ISSN: 0004-6256. DOI: 10.1086/508988.

Merten, J. et al. (June 2009). "Combining Weak and Strong Cluster Lensing: Applications to Simulations and MS 2137". In: *Astron. Astrophys.* 500, pp. 681–691. ISSN: 0004-6361. DOI: 10.1051/0004-6361/200810372.

Merten, Julian (Sept. 2016). "Mesh-Free Free-Form Lensing - I. Methodology and Application to Mass Reconstruction". In: *Monthly Notices of the Royal Astronomical Society* 461, pp. 2328–2345. ISSN: 0035-8711. DOI: 10.1093/mnras/stw1413.

Meyer, Sven, Matthias Redlich, and Matthias Bartelmann (Mar. 2015). "Evolution of Linear Perturbations in Lemaítre-Tolman-Bondi Void Models". In: *Journal of Cosmology and Astro-Particle Physics* 03, p. 053. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2015/03/053.

Mukhanov, V. F. and G. V. Chibisov (May 1981). "Quantum Fluctuations and a Nonsingular Universe". en. In: *Sov. J. Exp. Theor. Phys. Lett.* 33, p. 532.

Navarro, Julio F., Carlos S. Frenk, and Simon D. M. White (May 1996). "The Structure of Cold Dark Matter Halos". In: *Astrophys. J.* 462, p. 563. ISSN: 0004-637X. DOI: 10.1086/177173.

— (Dec. 1997). "A Universal Density Profile from Hierarchical Clustering". In: *The Astrophysical Journal* 490, pp. 493–508. ISSN: 0004-637X. DOI: 10.1086/304888.

Peebles, P. J. E. (1980). "The Large-Scale Structure of the Universe". en. In: *Large-Scale Struct. Universe Phillip James Edwin Peebles Princet. Univ. Press*.

Planck Collaboration et al. (Sept. 2016). "Planck 2015 Results. XIII. Cosmological Parameters". In: *Astronomy & Astrophysics, Volume 594, id.A13, 63 pp.* 594, A13. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201525830.

Pollina, Giorgia et al. (July 2017). "On the Linearity of Tracer Bias around Voids". In: *Monthly Notices of the Royal Astronomical Society* 469, pp. 787–799. ISSN: 0035-8711. DOI: 10.1093/mnras/stx785.

Postman, Marc et al. (Apr. 2012). "The Cluster Lensing and Supernova Survey with Hubble: An Overview". In: *The Astrophysical Journal Supplement Series* 199, p. 25. ISSN: 0067-0049. DOI: 10.1088/0067-0049/199/2/25.

Press, W. H. and P. Schechter (Feb. 1974). "Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation". In: *Astrophys J* 187, pp. 425–438.

Press, W. H. et al. (1992). *Numerical Recipes in C. The Art of Scientific Computing*. Second. Cambridge: University Press.

Ragagnin, A. et al. (July 2017). "A Web Portal for Hydrodynamical, Cosmological Simulations". In: *Astronomy and Computing* 20, pp. 52–67. ISSN: 2213-1337. DOI: 10.1016/j.ascom.2017.05.001.

Reblinsky, K. (Dec. 2000). "Cluster Deprojection Combining Multiple Observable Data Sets". In: *Astronomy and Astrophysics, v.364, p.377-390 (2000)* 364, pp. 377–390. ISSN: 0004-6361.

Rees, M. J. and D. W. Sciama (Feb. 1968). "Large-Scale Density Inhomogeneities in the Universe". en. In: *Nature* 217.5128, p. 511. DOI: 10.1038/217511a0.

Refregier, Alexandre (Jan. 2003). "Shapelets - I. A Method for Image Analysis". en. In: *Mon. Not. R. Astron. Soc.* 338.1, p. 35. DOI: 10.1046/j.1365-8711.2003.05901.x.

Reyes, Reinabelle et al. (Mar. 2010). "Confirmation of General Relativity on Large Scales from Weak Lensing and Galaxy Velocities". In: *Nature* 464, pp. 256–258. ISSN: 0028-0836. DOI: 10.1038/nature08857.

Robertson, H. P. (Nov. 1935). "Kinematics and World-Structure". en. In: *Astrophys. J.* 82, p. 284. DOI: 10.1086/143681.

Rubin, V. C., W. K. Ford, and N. Thonnard (Nov. 1978). "Extended Rotation Curves of High-Luminosity Spiral Galaxies. IV - Systematic Dynamical Properties, SA through SC". en. In: *Astrophys. J.* 225, pp. L107–L111. ISSN: 0004-637X. DOI: 10.1086/182804.

Sachs, R. K. and A. M. Wolfe (Jan. 1967). "Perturbations of a Cosmological Model and Angular Variations of the Microwave Background". en. In: *Astrophys. J.* 147, p. 73. DOI: 10.1086/148982.

Sarli, Eleonora et al. (Oct. 2014). "Reconstructing the Projected Gravitational Potential of Galaxy Clusters from Galaxy Kinematics". In: *Astron Astrophys* 570, A9. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201321748.

Schmid, C. et al. (July 2010). "Simulations of X-Ray Telescopes for eROSITA and IXO". In: *X-ray Astronomy 2009; Present Status, Multi-Wavelength Approach and Future Perspectives* 1248, pp. 591–592. ISSN: 0094-243X. DOI: 10.1063/1.3475349.

Schneider, Peter, Christopher Kochanek, and Joachim Wambsganss (2006). *Gravitational Lensing: Strong, Weak and Micro: Saas-Fee Advanced Course 33*. en. Ed. by Georges Meylan, Philippe Jetzer, and Pierre North. Saas-Fee Advanced Course. Berlin Heidelberg: Springer-Verlag. ISBN: 978-3-540-30309-1.

Schneider, Peter and Dominique Sluse (Apr. 2014). "Source-Position Transformation: An Approximate Invariance in Strong Gravitational Lensing". In: *Astronomy and Astrophysics* 564, A103. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201322106.

Schuecker, P. et al. (Feb. 2003). "The REFLEX Galaxy Cluster Survey. VII. $\mathrm{Omega_m}$ and $\mathrm{Sigma_8}$ from Cluster Abundance and Large-Scale Clustering". en. In: *Astron. Astrophys.* 398, p. 867. DOI: 10.1051/0004-6361:20021715.

Schwinn, J. et al. (Dec. 2018). "Uncovering Substructure with Wavelets: Proof of Concept Using Abell 2744". In: *Monthly Notices of the Royal Astronomical Society* 481, pp. 4300–4310. ISSN: 0035-8711. DOI: 10.1093/mnras/sty2566.

Seljak, Uroš (Oct. 2000). "Analytic Model for Galaxy and Dark Matter Clustering". In: *Monthly Notices of the Royal Astronomical Society* 318, pp. 203–213. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2000.03715.x.

Sereno, Mauro and Stefano Ettori (July 2017). "CoMaLit - V. Mass Forecasting with Proxies: Method and Application to Weak Lensing Calibrated Samples". en. In: *Mon. Not. R. Astron. Soc.* 468.3, p. 3322. DOI: 10.1093/mnras/stx576.

Shannon, C. E. (July 1948). "A Mathematical Theory of Communication". In: *Bell Syst. Tech. J.* 27.3, pp. 379–423. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Sheth, Ravi K., H. J. Mo, and Giuseppe Tormen (May 2001). "Ellipsoidal Collapse and an Improved Model for the Number and Spatial Distribution of Dark Matter Haloes". In: *Mon. Not. R. Astron. Soc.* 323, pp. 1–12. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2001.04006.x.

Shi, Xun and Eiichiro Komatsu (July 2014). "Analytical Model for Non-Thermal Pressure in Galaxy Clusters". In: *Mon. Not. R. Astron. Soc.* 442, pp. 521–532. ISSN: 0035-8711. DOI: 10.1093/mnras/stu858.

Skilling, John (Nov. 2004). "Nested Sampling". en. In: *Am. Inst. Phys. Conf. Ser.* 735, p. 395. DOI: 10.1063/1.1835238.

Spergel, David N. and Paul J. Steinhardt (Apr. 2000). "Observational Evidence for Self-Interacting Cold Dark Matter". en. In: *Phys. Rev. Lett.* 84.17, p. 3760. DOI: 10.1103/PhysRevLett.84.3760.

Springel, Volker (Dec. 2005). "The Cosmological Simulation Code GADGET-2". In: *Mon. Not. R. Astron. Soc.* 364, pp. 1105–1134. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09655.x.

— (Sept. 2010). "Smoothed Particle Hydrodynamics in Astrophysics". en. In: *Annu. Rev. Astron. Astrophys.* 48, p. 391. DOI: 10.1146/annurev-astro-081309-130914.

Springel, Volker et al. (Dec. 2001). "Populating a Cluster of Galaxies - I. Results at [Formmu2]Z=0". en. In: *Mon. Not. R. Astron. Soc.* 328.3, p. 726. DOI: 10.1046/j.1365-8711.2001.04912.x.

Subramanian, K. and S. A. Cowling (Mar. 1986). "On Local Conditions for Multiple Imaging by Bounded, Smooth Gravitational Lenses". In: *Mon R Astron Soc* 219, pp. 333–346.

Sunyaev, R. A. and I. B. Zeldovich (1980). "Microwave Background Radiation as a Probe of the Contemporary Structure and History of the Universe". In: *Annu Rev Astron Astrophys* 18, pp. 537–560.

Suto, Yasushi, Shin Sasaki, and Nobuyoshi Makino (Dec. 1998). "Gas Density and X-Ray Surface Brightness Profiles of Clusters of Galaxies from Dark Matter Halo Potentials: Beyond the Isothermal β-Model". In: *Astrophys. J.* 509. DOI: 10.1086/306520.

Suyu, S. H. et al. (Sept. 2006). "A Bayesian Analysis of Regularized Source Inversions in Gravitational Lensing". In: *Mon. Not. R. Astron. Soc.* 371, pp. 983–998. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2006.10733.x.

Suyu, S. H. et al. (July 2017). "H0LiCOW - I. H0 Lenses in COSMOGRAIL's Wellspring: Program Overview". In: *Monthly Notices of the Royal Astronomical Society* 468, pp. 2590–2604. ISSN: 0035-8711. DOI: 10.1093/mnras/stx483.

Tagore, Amitpal S. and Charles R. Keeton (Nov. 2014). "Statistical and Systematic Uncertainties in Pixel-Based Source Reconstruction Algorithms for Gravitational Lensing". In: *Mon. Not. R. Astron. Soc.* 445, pp. 694–710. ISSN: 0035-8711. DOI: 10.1093/mnras/stu1671.

Tchernin, C. et al. (June 2018). "Reconstruction of the Two-Dimensional Gravitational Potential of Galaxy Clusters from X-Ray and Sunyaev-Zel'dovich

Measurements". In: *Astronomy and Astrophysics* 614, A38. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201629364.

Tikhonov, A. N. et al. (1995). *Numerical Methods for the Solution of Ill-Posed Problems*. en. Mathematics and Its Applications. Springer Netherlands. ISBN: 978-0-7923-3583-2.

Tinker, Jeremy et al. (Dec. 2008). "Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality". In: *The Astrophysical Journal* 688, pp. 709–728. ISSN: 0004-637X. DOI: 10.1086/591439.

Tormen, Giuseppe, Francois R. Bouchet, and Simon D. M. White (Apr. 1997). "The Structure and Dynamical Evolution of Dark Matter Haloes". In: *Mon. Not. R. Astron. Soc.* 286, pp. 865–884. ISSN: 0035-8711.

Trenti, Michele and Piet Hut (May 2008). "N-Body Simulations (Gravitational)". en. In: *Scholarpedia* 3.5, p. 3930. ISSN: 1941-6016. DOI: 10.4249/scholarpedia.3930.

Trümper, Joachim E. and Günther Hasinger (2008). "The Universe in X-Rays". en. In: *Universe X-Rays*. DOI: 10.1007/978-3-540-34412-4.

Tugendhat, Tim M. and Björn Malte Schäfer (May 2018). "Angular Ellipticity Correlations in a Composite Alignment Model for Elliptical and Spiral Galaxies and Inference from Weak Lensing". In: *Monthly Notices of the Royal Astronomical Society* 476, pp. 3460–3477. ISSN: 0035-8711. DOI: 10.1093/mnras/sty323.

Vegetti, S. and L. V. E. Koopmans (Jan. 2009). "Bayesian Strong Gravitational-Lens Modelling on Adaptive Grids: Objective Detection of Mass Substructure in Galaxies". In: *Mon. Not. R. Astron. Soc.* 392, pp. 945–963. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2008.14005.x.

Vegetti, S. et al. (Jan. 2012). "Gravitational Detection of a Low-Mass Dark Satellite Galaxy at Cosmological Distance". In: *Nature* 481, pp. 341–343. ISSN: 0028-0836. DOI: 10.1038/nature10669.

Visinelli, Luca and Paolo Gondolo (Aug. 2009). "Dark Matter Axions Revisited". en. In: *Phys. Rev. D* 80.3, p. 035024. DOI: 10.1103/PhysRevD.80.035024.

Voit, G. Mark (Apr. 2005). "Tracing Cosmic Evolution with Clusters of Galaxies". In: *Rev. Mod. Phys.* 77, pp. 207–258. ISSN: 0034-6861. DOI: 10.1103/RevModPhys.77.207.

Voit, G. Mark and Greg L. Bryan (Nov. 2001). "Regulation of the X-Ray Luminosity of Clusters of Galaxies by Cooling and Supernova Feedback". en. In: *Nature* 414.6862, p. 425. DOI: 10.1038/35106523.

Walker, A. G. (1937). "On Milne's Theory of World-Structure". en. In: *Proc. Lond. Math. Soc.* 42, p. 90. DOI: 10.1112/plms/s2-42.1.90.

Warren, S. J. and S. Dye (June 2003). "Semilinear Gravitational Lens Inversion". In: *The Astrophysical Journal* 590, pp. 673–682. ISSN: 0004-637X. DOI: 10.1086/375132.

Weinberg, Steven (Jan. 1978). "A New Light Boson?" en. In: *Phys. Rev. Lett.* 40.4, p. 223. DOI: 10.1103/PhysRevLett.40.223.

— (2008). *Cosmology*. Oxford University Press. ISBN: 978-0-19-852682-7.

White, S. D. M., C. S. Frenk, and M. Davis (Nov. 1983). "Clustering in a Neutrino-Dominated Universe". In: *The Astrophysical Journal Letters* 274, pp. L1–L5. ISSN: 0004-637X. DOI: 10.1086/184139.

Zhao, Gong-Bo et al. (Dec. 2009). "Cosmological Tests of General Relativity with Future Tomographic Surveys". en. In: *Phys. Rev. Lett.* 103.24, p. 241301. DOI: 10.1103/PhysRevLett.103.241301.

Zitrin, Adi and Tom Broadhurst (Oct. 2009). "Discovery of the Largest Known Lensed Images Formed by a Critically Convergent Lensing Cluster". en. In: *Astrophys. J.* 703.2, p. L132. DOI: 10.1088/0004-637X/703/2/L132.

Zwicky, F. (1933). "Die Rotverschiebung von Extragalaktischen Nebeln". In: *Helvetica Physica Acta* 6, pp. 110–127. ISSN: 0018-0238.

Acknowledgements

That's it, folks, I'm done. It was (mostly) fun while it lasted, but I'm (mostly) glad that it's over.

I wish to thank Stefan Hilbert for giving me the opportunity to do this project and for having far more patience with me than I could have reasonably expected. My thanks also go to Andreas Burkert for taking on the job as my official supervisor.

I want to thank Jochen Weller for granting me refuge at his group every Friday, which allowed me to participate in fun discussions and shortened my commute at least once a week. All the members of the physical cosmology group at USM have been most welcoming and I enjoyed my visits there very much. I thank Kerstin Paech for lots of scientific insights, mockery and support, Tea, Giorgia and Steffen for sharing their precious office space with me and for lots and lots of fun conversations, physics discussions of varying obscurity, movie nights and assorted shenanigans... and of course everyone else for making lunch a blast.

Lunch brings us back to Garching, where our lunch group provided the much appreciated, sometimes heated discussions of random topics which seem to be, at least for me, the backbone of an academic career. Special thanks in this context go to Haakon and Damien for providing many of said topics and of course Graham for being even more of a grump than myself (and also enabling me to waste time at coffee, at table tennis, etc.). I thank Simona Vegetti for a lot of insightful discussions (it's how I learn), be it on physics or politics, and I would like to apologise here for not finishing the lensing project we started together.

Outside the Munich astrophysics circles, I express gratitude to my long-time collaborators Celine Tchernin and Julian Merten, to Matthias Bartelmann, who got me started on this whole business in the first place, and to Björn Schäfer, Tim Tugendhat and the nice folks at ITA, who made my stays there always fun. I want to especially thank Elena Kozlikin and Robert Lilow for organising the amazing Schöntal workshops and letting me participate in them.

Finally I thank Steffen for proofreading many parts of this thesis, which helped me a lot.

Now, I obviously didn't mention everyone who'd deserve it and I clearly only mentioned people I met through my astrophysical endeavours (I want this to fit on one page). For the rest, suffice it to say that life has so far always dealt me very agreeable sets of humans wherever I went and that I appreciate that fact quite a bit.