The spatial distribution of dark matter in the Universe

Tamina Lund

Lund Observatory
Lund University

Abstract

It was previously found that the density profile of dark matter halos is well-fit by a universal, smooth function over a vast range of masses and redshifts, named the Navarro-Frenk-White (NFW) profile after its discoverers (Navarro et al. 1996, 1997). Despite its impressive goodness of fit, the NFW model remains an approximation of the density profile of realistic halos. In the literature, this model has been used for fitting the density profiles across a cosmological population of halos. Deriving the best-fit parameters using such an approach enables us to establish a correlation between halo mass and concentration (e.g. Dutton & Macciò 2014). However, this classical approach neglects uncertainties associated with the construction of a density profile from simulation data. Therefore, this thesis aims to quantify how such uncertainties propagate into our inference of the best-fit parameters. For a halo density profile gained from simulation data, point estimates for the best-fit parameters were gained from the method of least squares, before estimating the full probability distribution thereof using Bayesian inference. From the application of a physical prior in the latter method, it was found that despite having taken Poisson noise into account, the uncertainties had been strongly underestimated, as the best-fit parameters from the prior were far outside the confidence levels of the likelihood contour. A further point underlining this was found when applying the literature method to the halo treated in this project, whose best-fit parameters were found to not be compatible on a 10 sigma level with the ones from least squares. This suggests that uncertainties in the data forming the halo density profile are dominated by a systematic source of error, outweighing the contribution arising from the Poisson noise. Modelling and quantifying such systematic uncertainties will be key for future studies aiming to accurately pinpoint the correlation between halo mass and concentration.

Populärvetenskaplig beskrivning

Mörk materia är en av huvudkomponenterna av universum. Det går inte att observera den på något direkt sätt, men dess närvaro är uppenbar, eftersom rörelsen av stjärnor och gas påverkas av dess dragningskraft. Mörk materia bildar gravitationsbundna klumpar som kallas för halos. Att förstå dessa objekt utgör en grundläggande del av kosmologin, vilket är vetenskapen som handlar om universums usprung och utveckling, såväl som dess storskaliga strukturer.

Ett vanligt tillvägagångssätt inom all slags vetenskap är att antingen göra observationer direkt eller att återge ett fenomen från verkligheten så noggrant som möjligt genom simuleringar och att sen jämföra denna data med en modell, för att se hur bra de stämmer överens. En modell är en funktion som antas ha ett liknande beteende som datan i fråga och innehåller vanligtvis en eller flera parametrar. Målet är att få modellen att sammanfalla med datan så bra som möjligt genom att hitta lämpliga värden för parametrarna. Det kan bland annat användas till att undersöka hur väl modellen faktiskt beskriver datan.

För visualiseringen av fördelningen av mörk materia i en halo är det lämpligt att skapa densitetsprofiler. Dessa ges av en funktion som beskriver densiteten vid en given radie i halon. Det har tidigare visat sig att densitetsprofilerna för alla halos följer samma form, benämnda Navarro-Frenk-White (NFW) profil efter dess upptäckare (Navarro et al. 1997). Eftersom alla halos kan beskrivas av detta uttryck, så är det meningsfullt att använda det som modell, vilket gjordes i detta projekt. Vidare, så hittades lämpliga värden för dess parametrar som fick modellen att sammanfalla med datan, i antagandet att modellen var bra. En viktig poäng är att osäkerheterna i varje datapunkt togs i beräkning, vilket har en påverkan på kurvanpassningen av modellen till datan. Osäkerheter innebär att en punkt inte bara ges av ett enskilt värde, utan är definierad över ett intervall. Eftersom detta inte hade gjorts vid tidigare tillfällen, så var målet att ta reda på hur det skulle påverka parametrarna och deras osäkerheter och att därmed undersöka om det är giltigt att försumma datapunkternas osäkerheter eller inte.

Contents

List of Figures

Chapter 1

Introduction

The constituents of the Universe can be divided into dark energy, dark matter and normal matter, where dark matter makes up as much as 27% of today's Universe (Planck Collaboration et al. 2018). Despite not being directly detectable, dark matter affects the dynamics of stars and gas in galaxies, which is most easily seen in rotation curves of the Milky Way or through the observation of gravitational lensing (Clowe et al. 2006). Through gravity, dark matter forms gravitationally bound clumps, which surround all galaxies (Wechsler & Tinker 2018) and bear the name dark matter halos. Often, there are smaller halos within such dark matter halos, which are called subhalos. Halos create a fundamental unit in cosmology, their understanding being crucial in research areas such as galaxy formation.

In the early Universe, immediately after the Big Bang, there was a period of inflation, in which density fluctuations were created. These perturbations grew linearly with the continued expansion of the Universe, until a critical density was reached. Consequently, they collapsed and formed halos, which grew by accretion of matter from their surroundings and through mergers with other halos (Houjun Mo & White 2010).

Density profiles of dark matter halos, which is the main focus of this project, are a physical representation of the radial distribution of dark matter. Knowledge thereof helps to gain a deeper understanding of a halo's structure, as well as being one of many components leading to a better understanding of dark matter as a whole, as this is yet a domain leaving researchers with many questions. A common assumption when making such density profiles is to treat the halos as spherically symmetric. This greatly simplifies the process of making a density profile, as well as any calculations done in connection with it.

Previously, the surprising observation was made that the majority of all dark matter halos in equilibrium can be fit by the empirical expression for the density shown in equation (1.1) (Navarro et al. 1997). This was found to hold true for a vast range of halo masses and redshifts (Jing 2000) and implies that the shape of halo density profiles is universal. The profile was named NFW profile after its discoverers Navarro, Frenk and White and an example thereof is shown in figure 1.1.
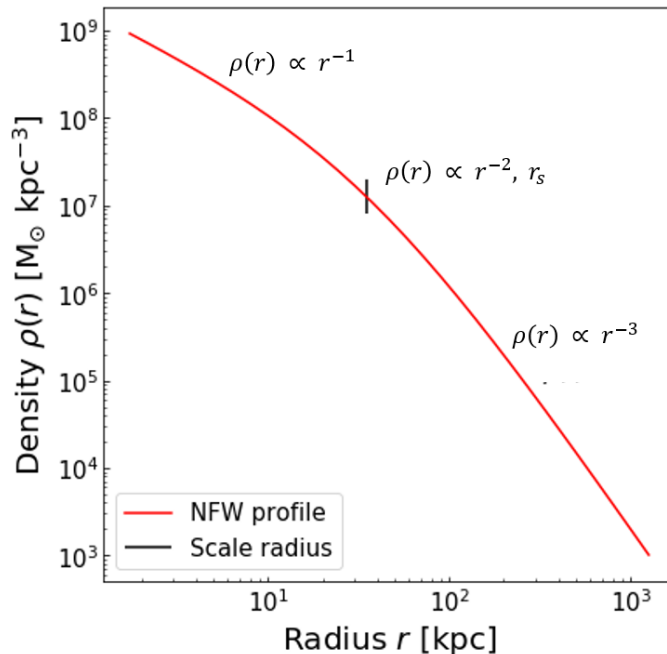
Figure 1.1: The NFW profile of a dark matter halo shows the spatial distribution of its density with respect to radius. It is a function of two parameters and shows a divergence as the radius approaches zero. As can be seen in the figure, the profile has two regions of different slopes, where the transition is given by the scale radius.

$$\rho(r) = \frac{\rho_0}{\frac{r}{r_s}\left(1 + \frac{r}{r_s}\right)^2}, \tag{1.1}$$

where $\rho(r)$ is the density of dark matter at a certain radius and ρ_0 is a normalisation factor. A halo's radial position in spherical coordinates is given by r, while r_s is called scale radius. ρ_0 and r_s are parameters, which vary from halo to halo. With regard to parameters, an important note on the topic of degeneracies can be made. If parameters are degenerate, this means that different sets of parameter values give the same profile for one specific halo and hence, that the density profile is not uniquely determined by its parameters. While the parameters are not degenerate in the theoretical model, degeneracies can occur in practical cases where the model is fit to incomplete data, meaning that uncertainties in the data allow two pairs of parameters to fit equally well.

As shown in figure 1.1, the density profile of a dark matter halo can be divided into two regions, separated by the scale radius. The region on the left, describing the inner parts of a halo, ranges from the centre of the halo to the scale radius. It has a shallow slope described by $\rho(r) \propto r^{-1}$, while the second region ranges between the scale radius and the outer edge of the halo. The latter region of the profile shows a considerably steeper slope of $\rho \propto r^{-3}$. In order to gain the renowned shape of the profile, both axes must be in logarithmic scale. This is also advantageous, given the fact that the profile usually ranges over several orders of magnitude. Equation (1.1) says that with an increasing scale radius

or in other words, the further out the transition from the inner to the outer region in the density profile is for a fixed ρ_0, the more the denominator decreases and the smaller the density in the halo centre becomes. This implies that halos with smaller scale radii have larger central densities.

Additionally, it is important to emphasize that the outer edge of a halo is not formerly identifiable, but rather, is chosen to be equal to the halo's virial radius. The virial radius is often written as r_{200} and describes the radius at which the density of the halo is 200 times greater than the critical density of the Universe at the halo's redshift (Navarro 1995, Navarro et al. 1996).

In the past, a common approach to finding the best-fit NFW model for a halo involved finding values for the parameters r_s and ρ_0 that minimize ψ^2 in equation (1.2), adapted from Ludlow et al. (2016)

$$\psi^2 = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} [\ln \rho_i - \ln \rho_{NFW}(\rho_0; r_s)]^2, \tag{1.2}$$

where N_{bin} is the total number of bins, ρ_i is the density of bin i, given by the data from a simulation and finally, $\rho_{NFW}(\rho_0; r_s)$ is the value for the halo density, given by the fitted model at that point. This fitting method is similar to the least squares fitting method, which will be used in this project and is explained in more detail in Chapter 2.2. The two methods are similar, because both describe a minimization problem, where the gradient is calculated in an iterative process for different sets of parameter values, until a minimum for in this case ψ^2 is found. In the literature method, the obtained value for r_s was then used to find the concentration c of the halo in question by using equation (1.3) (Ludlow et al. 2016).

$$c \equiv \frac{r_{200}}{r_s} \tag{1.3}$$

The application of this procedure to every halo within a cosmological volume then recovered the correlation between halo mass and concentration, with halos of higher mass having lower concentrations (Ludlow et al. 2016).

However, this procedure lacks in clarity with regard to the uncertainties of the data. First of all, these uncertainties were not taken into account, meaning that when using equation (1.2) to find the best fit between model and data, each data point was given equal weight. Consequently, the found best-fit parameter values were merely point estimates assuming that the data was perfectly known, rather than probability distributions. As a result, the use of the obtained values for r_s when finding the concentrations lead to the latter also becoming point estimates. Furthermore, it is important to underline that while the NFW model is a suitable model for fitting halo density profiles, its goodness of fit varies for different halos. For these reasons, there is a prevailing unclarity over how such uncertainties propagate in the inference of the mass-concentration relationship, which is a key input for galaxy formation models (e.g. Somerville et al. 2008).

Due to the lack of clarity over how much uncertainty the classical approach to fitting the

NFW profile to halo density profiles has, this project aims to improve the methods to understand whether or not the assumption of negligible uncertainties for the parameters is indeed valid. This was done by constructing a halo density profile based on simulation data, described in chapter 2.1, followed by assigning uncertainties to the data points assuming Poisson distributions. The density profile was then fit using the method of least squares, in order to find the set of parameters that gives the best fit between the NFW model and the data. As least squares only gives point estimates for the parameters, the full probability distribution function of the model parameters was in a later step estimated using Bayesian inference, which is described in section 2.2.3. While initially using a flat prior for this, a physical prior was used at a later point, which was found based on previous knowledge about the mass-concentration relation between dark matter halos. Finally, the thesis closes with a comparison between the literature method applied to the halo density profile treated in this project and the least squares fit thereof, where only the latter takes uncertainties of the data points into account.

Chapter 2

Methods

2.1 Obtaining a halo density profile from simulation data

The first task of this project consisted of using data from an N-body simulation of dark matter halos, in order to retrieve a density profile of one of them, using the programming language Python. This halo was found to have virial mass $M_{200} = 1.38 \cdot 10^{13}$ $\mathrm{M}_\odot$ and consisted of 14340 particles. The simulation that was used is a cosmological simulation of structure formation (Rey et al. 2019), which exclusively models the evolution of dark matter, ignoring usual matters constituting for example gas and dust. It discretizes space into collisionless particles of equal mass, whose motions depend on interactions with their surrounding particles, following Newtonian gravity (Houjun Mo & White 2010). The simulation was run using the code GADGET-2 (Springel 2005) and was run to be consistent with cosmological parameters obtained by the Planck satellite. The obtained density profile was subsequently compared to one gained using built-in functions from the analysis package Pynbody (Pontzen et al. 2013), which is a common tool for the manipulation of N-body simulations in astronomy. The comparison was done for the sake of verification of the code, as the two density profiles should coincide. The following sections give an explanation of each step underlying the process of retrieving a halo density profile.

In Pynbody, the halos of a simulation are arranged in descending order with respect to mass. Arbitrarily, halo 50 was chosen to be used for this project. With Pynbody, the centre, meaning $(x, y, z) = (0, 0, 0)$ in the underlying coordinate system, was chosen to coincide with the mass centre of halo 50. For simplification purposes, the assumption was made that the halo is spherically symmetric. In a next step, the distance between the centre of the halo and each particle within it was found, prior to dividing the halo into logarithmically spaced shells. A logarithmic spacing was chosen, because the data spreads across several orders of magnitude.

Each shell corresponds to a bin, meaning that each of these previously mentioned shells that fill up the whole volume of the halo contains a certain number of particles. The bin size was chosen small enough that the assumption that all particles within a bin are at

equal distance from the halo's center is valid, but large enough to not unnecessarily increase the noise in the profile. The terms "bin" and "shell" can be used interchangeably here and in this project, the former will be used from now on. The number of particles within each bin was found by investigating whether or not each of the previously found distances lies within the range of $[r, r + \Delta r]$ for each bin respectively. The volume V_i of each bin was calculated using the approximated expression for the volume of a spherical shell, based on the assumption that the increment is much less than the radius of the halo ($\Delta r << r$). It is given by equation (2.1); its derivation can be found in appendix A. Using the approximated expression for the volume of a shell proved to be sufficient for this task, as the obtained profile and that retrieved with Pynbody, where the latter uses the exact expression, were found to be very similar. Additionally, the choice of expression for the volume does not have a strong impact on the final results.

$$V_i = 4\pi r_i^2 \Delta r \qquad (2.1)$$

From the number of particles N_i in each bin and the fact that each particle in the halo has equal mass $m_{particle}$, the total mass of each bin was found using equation (2.2).

$$m_i = N_i \cdot m_{particle}, \qquad (2.2)$$

The final step consisted of calculating the density ρ_i of each bin using equation (2.3) and plotting this with respect to the radius to obtain a density profile.

$$\rho_i = \frac{m_i}{V_i} \qquad (2.3)$$

2.2 Fitting the NFW profile

After having obtained a halo density profile based on data from an N-body simulation, the next step consisted of assigning uncertainties to each data point, before fitting the NFW model, given by equation (1.1), to the profile. The best-fit parameters were found using the least squares approach. In a later step, the full probability distribution function of the model parameters was estimated using Bayesian inference, both for a flat and a physical prior.

2.2.1 Assigning data uncertainties

Prior to making a fit of the data, each data point was assigned an uncertainty with regard to its density, represented as an error bar in y-direction. Taking uncertainties into account is crucial for curve fitting and the inherently linked parameter estimation, as these play the role of weighting the bins. Specifically, the uncertainties are chosen in a way that bins with a smaller number of particles are given a larger uncertainty and hence, less weight

than ones with a larger number of particles. As bins in the inner parts of halos tend to have smaller numbers of particles, they are given a larger error bar, implying a larger value for σ, which is the symbol used for the uncertainty. The latter is commonly referred to as the standard deviation of a distribution. The expression for σ varies from problem to problem, since it depends on which kind of distribution the errors follow, as well as having to match the units of the variable that it is an uncertainty of. In the problem that this project is faced with, it was assumed that the error bar in each bin follows the statistics of its number count, meaning that Poisson errors are the dominant source of error. Moreover, the assumption was made that bins are uncorrelated with one another. While the Poisson distribution is skewed for a small number of counts, the approximation of describing the errors by a Gaussian distribution becomes gradually more valid with an increasing number of particles per bin (Fageot et al. 2017). Therefore, the assumption was made that the counts per bin are large enough to assume the Poisson distribution to be symmetrical. Based on the above assumptions, an expression for the uncertainties of the densities of the data points was found, which is given by equation (2.4). A derivation thereof can be found in appendix B.

$$\sigma_i = \frac{1}{\sqrt{N_i}} \frac{m_i}{V_i},\tag{2.4}$$

While the first term given by $\frac{1}{\sqrt{N_i}}$ represents the general expression for the standard deviation of a Poisson distribution, it was multiplied by m_i/V_i to get suitable units for σ_i, matching those of the density. Taking the inverse of $\sqrt{N_i}$ has the effect of giving more weight and hence a smaller σ to bins containing a larger number of particles, as previously explained.

2.2.2   Finding the maximum likelihood with least squares

Least squares is a common fitting method and has the goal of finding the set of parameter values that gives the best fit between model and data. It does so by producing a fit that minimizes the sum of the square of all individual distances between each data point and the fitted curve, which is called chi squared and is given by equation (2.5) (Sivia & Skilling 2006).

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f_i(x)}{\sigma_i} \right)^2,\tag{2.5}$$

where y_i are the data points, $f_i(x)$ are the values gained from the model and σ_i is the uncertainty of each bin, as was explained in the preceding section. The term in the numerator is commonly referred to as residual, being the distance between each data point and the fit at that point. Figure 2.1 serves as an illustration of how the distances in the least squares fitting method are determined, in particular showing the individual distances between the

fit and the data points, the sum of the square of which is to be minimized. Equation 2.5 is a special case of the more general form of least squares, as it minimizes the square of this sum by weighting each individual distance by its error bar as opposed to giving each data point equal weight. In both cases, it describes an optimization problem, where the gradient is calculated in an iterative process for a range of different parameters, as given by equation (2.6), until a minimum for χ^2 is found.

$$\frac{\partial \chi^2}{\partial r_s} = 0, \quad \frac{\partial \chi^2}{\partial \rho_0} = 0 \qquad (2.6)$$

It holds that the smaller the value of χ^2, the better the fit. The parameters that minimize χ^2 are by definition the ones with the maximum likelihood for the model to fit the data.
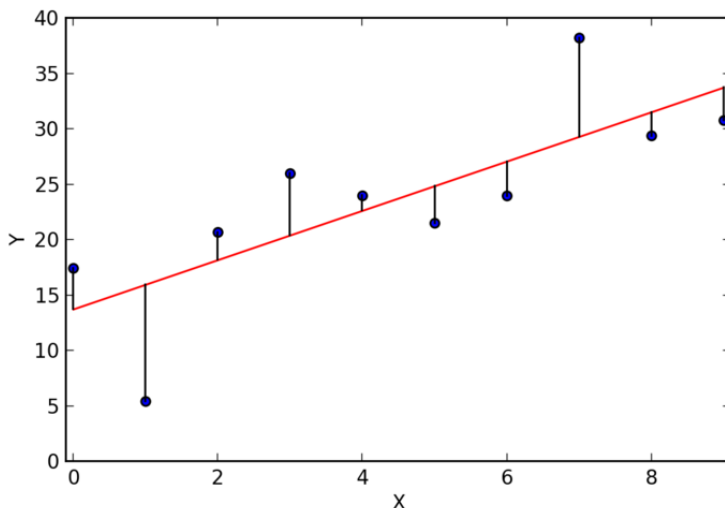

Figure 2.1: In the least squares fitting method the sum of the squared distances between the fit and each data point is minimized. In this figure, the fit is shown as a line passing through the data points, while the residual is given by the vertical lines connecting the points to the fit. (Haslwanter 2013)

Finally, it is worth mentioning that while least squares can be used to obtain parameter values that give the best fit of the model to the data, this method merely gives single values without uncertainties. This could mean that in cases where the probability distribution of one of the parameters has several similarly sizes peaks, every peak with exception of the highest is neglected. This implies that by solely applying least squares, no knowledge can be obtained about the existence of further peaks. For this reason, it makes sense to apply a different method, which calculates the likelihood for a vast range of parameter values and to in that way explore parameter space, in order to on the one hand verify that no peaks in the probability distributions of the parameters have been missed. On the other hand, this can be useful for the investigation of the existence of parameter degeneracies. Even more importantly, this method has the benefit that the resulting set of best-fit parameters

is not merely given by a set of single values, but rather by probability distributions of a certain width, giving information about the uncertainties. This process can either be done by explicitly calculating the likelihood for each combination of parameter values in a fixed parameter space or by applying an algorithm, where the former alternative was chosen for this project. For the calculation of the likelihood, the assumption was made that it is proportional to χ^2. This is possible when the likelihood can be assumed to have Gaussian distributions, which is valid when the number of particles is large (Davis & Parkinson 2016). Hence, for each combination of parameter values in the parameter grid, the value of χ^2 was calculated, from which the likelihood $\mathcal{L}$ was calculated using equation (2.7) (Davis & Parkinson 2016).

$$\mathcal{L} \propto \exp(-0.5 \cdot \chi^2) \tag{2.7}$$

2.2.3 Finding a posterior for the parameters with Bayesian inference

There exist two fundamentally different approaches when dealing with problems of statistical nature, which are commonly divided into the frequentist and the Bayesian approach. In the frequentist approach, an experiment is executed multiple times, where each outcome may differ slightly from the previous, due to systematical errors. It then takes the probability to be a direct consequence of the outcomes of the experiments, which is commonly used in the laboratory (Riemer-Sørensen 2017). This implies that the data is treated probabilistically, as it changes with each iteration of the experiment, while the parameters are fixed (Leistedt 2017). In Bayesian inference on the other hand, the experiment is only executed once, but the resulting probability of the model being correct for a given set of parameter values can be updated once further knowledge of the outcome becomes available. This probability is called posterior and is given by Bayes' theorem, which is stated in equation (2.8) (Sivia & Skilling 2006) and involves including prior knowledge about the result. Conversely to the frequentist approach, the data remains unchanged, while the model parameters are treated probabilistically (Leistedt 2017). In cosmology, it can be favourable to apply Bayesian inference when dealing with data analysis, because there is only one universe, meaning that an experiment can be executed only once. However, the knowledge about it can be updated when using this approach.

$$\mathrm{prob}(X|Y,I) = \frac{\mathrm{prob}(Y|X,I)\,\mathrm{prob}(X|I)}{\mathrm{prob}(Y|I)}, \tag{2.8}$$

where X stands for a set of parameters, Y is the data and I describes any external information, such as the physical knowledge that the model is based on (Riemer-Sørensen 2017). The different terms in this expression are given by $\mathrm{prob}(X|Y,I)$, which stands for the posterior, $\mathrm{prob}(Y|X,I)$, which is called likelihood, $\mathrm{prob}(X|I)$, which stands for the prior and finally, $\mathrm{prob}(Y|I)$, which is the evidence. The latter term is merely used for normalization purposes and is commonly disregarded unless it is of importance for the

problem. Unlike for model selection, knowledge thereof is unnecessary for the process of parameter estimation, as it is only dependent of the model, but not of the parameters (Riemer-Sørensen 2017). Hence, it was omitted in this project, as a fixed model was used. Doing so allows rewriting Bayes' theorem as follows (Sivia & Skilling 2006).

$$\text{prob}(X|Y, I) \propto \text{prob}(Y|X, I)\,\text{prob}(X|I) \tag{2.9}$$

Each of the terms represent probability distributions, meaning that the resulting posterior is not just given by a single value, but includes information about the uncertainty of the outcome as well. The following sections will give a more detailed explanation of the different terms in Bayes' theorem, followed by a discourse about marginalization. They are adjusted to the problem of this project, which is two-dimensional and deals with finding suitable model parameter values, using the NFW profile as a fixed model.

Likelihood

The likelihood describes the probability of the model being correct for a set of parameters, given the data. It is calculated from the experiment and has some uncertainty given by the uncertainties of the data points. The likelihood corresponds to the likelihood given by the frequentist approach, which can be found through application of optimization procedures such as least squares.

Prior

Bayesian inference entails the possibility of using previous knowledge about the model parameters before having seen the specific data of the problem (Sivia & Skilling 2006), which is done through the use of a prior. There exist different types of priors, which can be divided into the subcategories uninformative and informative prior respectively. The former is often referred to as flat prior and is used whenever there is only a minimum amount of information available about the model parameter values. As the probability distributions of the parameters can be assumed to be continuous, equal probability can be assigned to a certain range of values $[x_{min}, x_{max}]$, which is given by equation (2.10). The choice of interval is based on the prevailing knowledge about the parameters. Zero probability is assigned to the values outside of this range.

$$f(x) = \frac{1}{x_{max} - x_{min}}, \tag{2.10}$$

where $f(x)$ is the value assigned to the range $[x_{max}, x_{min}]$ and x_{min} and x_{max} are the smallest and largest values respectively within this interval.
Alternatively to a flat prior, an informative prior can be chosen, also referred to as physical prior. The application thereof is beneficial in cases where some knowledge about the parameter values is available, often taken from other experiments that give an understanding of the usual range of values within which the parameters fall, which sets constraints on

them. In that case, the prior can most easily be assumed to follow a Gaussian distribution, which is characterized by a mean and a width, the latter describing the uncertainty. The aim of applying a prior is to achieve a reduction of the uncertainty in the final posterior distribution, which is illustrated in figure 2.2. The idea of including a prior lies in the fact that the maximum posterior might differ from the maximum likelihood.
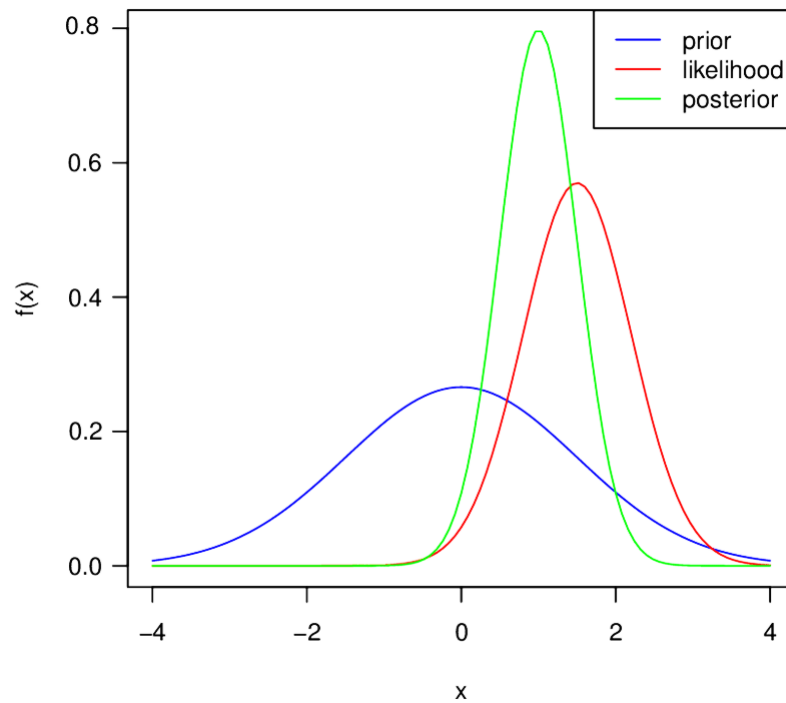

Figure 2.2: When applying Bayesian inference, prior information about the variables is multiplied by the likelihood, resulting in a posterior distribution. By doing so, the aim is to achieve a reduction of the uncertainty in the final distribution, which the figure shows an example of. (Aksu 2018)

Posterior

As could be seen from equation (2.9), the multiplication of the likelihood and the prior gives the posterior. As previously mentioned, it thus describes the probability of the model being correct for a given set of parameter values and can be updated once further knowledge of the outcome becomes available. Its uncertainty arises from the error bars embedded in both the likelihood and the prior respectively, as the likelihood has uncertainties originating from the error bars of the individual data points, while the prior has more trivial uncertainties, given by the width of its distribution. The posterior can be said to be similar to the likelihood, with the exception that prior knowledge has been folded in. It is of interest to investigate how much these uncertainties are reduced or in other words, how much narrower the posterior gets compared to the likelihood. If no reduction can be observed at all, there was nothing gained from the application of Bayesian inference.

Marginalization

As this project uses a model that has two model parameters, the prior, likelihood and posterior all become two-dimensional probability distributions, visualized as contours. For the sake of being able to compare the one-dimensional probability distributions of the prior, likelihood and posterior for each parameter separately, or simply in order to parameter-wise visualize the change of the median and uncertainty of a parameter with the aim of illustrating how the probability was affected by the prior, the process called marginalization can be applied. It extracts the one-dimensional probability distributions from the two-dimensional contour for each parameter individually and does so by compressing the contour to each axis through the application of equation (2.11) (Sivia & Skilling 2006), where the parameters ρ_0 and r_s of the NFW profile have been introduced already. This gives the probability distribution of one parameter marginalized over the whole likelihood, whose peak represents the parameter value of largest likelihood and the width describes its uncertainty on the likelihood level.

$$
\begin{aligned}
P(r_s) &= \int_0^\infty P(r_s, \rho_0) \; d\rho_0, \\
P(\rho_0) &= \int_0^\infty P(\rho_0, r_s) \; dr_s
\end{aligned}
\tag{2.11}
$$

As can be seen, marginalization describes the integration of the joint probability distribution with respect to the unwanted variable, which in this case is the other parameter. This needs to be done for both parameters separately. While marginalization can be used for the reduction of a two-dimensional probability distribution into one dimension, it in more general terms has the aim of removing any unwanted variables. This makes it not only a useful tool for retrieving one-dimensional probability distributions of parameters from two-dimensional contours, but also for removing nuisance parameters such as instrumental effects that are not of interest to the result, but need to be accounted for.

Chapter 3

Results

3.1 Halo density profile

A dark matter halo density profile, which is shown in figure 3.1, was retrieved based on the method described in chapter 2.2, using data from a cosmological N-body simulation of the structure formation of dark matter. For the sake of verification of the analysis code developed for this project, this profile was plotted against one gained from built-in functions in Pynbody. In the retrieved density profile, the two regions of the NFW profile are not as clearly visible as in the example shown in figure 1.1, due to r_s being at a small radius compared to the halo's total radius. Some noise can be seen in the inner parts of the halo, which is numerical noise due to a small number of particles in the innermost bins.

3.2 Least squares fit

After having obtained a density profile that was close to identical to the one obtained from Pynbody, a curve was fit to the data points constituting the profile, using the NFW profile as a model and the fitting method least squares. This was done in order to find the best-fit parameters. The first attempt was made without using any initial guesses for the parameters, meaning that Python used the default values of $(r_s, \rho_0) = (1, 1)$. The fit gained from this was undeniably bad, as it almost did not follow the data points at all, especially at smaller radii. Therefore, a new fit was made by introducing initial guesses for the parameters. These guesses were found by plotting the NFW function against the data and by trial and error finding parameters that made the model coincide with the data as well as possible. While approximate values for the initial guesses of the parameters could be found in this way, the fit remained deviant, due to the lack of error bars. Without including error bars, each bin was given equal weight, irrespective of the number of particles in a bin. While the inner bins only contained very few particles such as 1 or 2, the number gradually increased with radius. This means that too much weight had been given to the innermost bins when neglecting uncertainties. The problem of weighting was resolved by introducing Poisson error bars for the density, given by equation (2.4), which assigned a
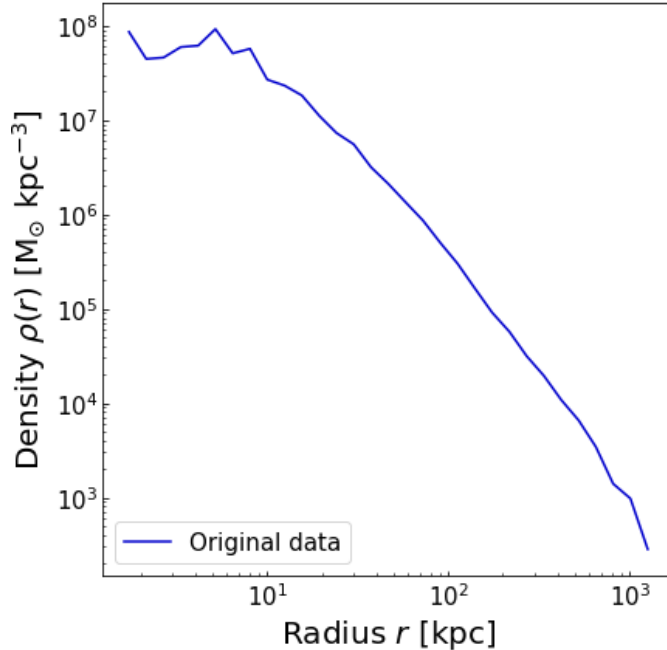
Figure 3.1: The figure shows the obtained halo density profile of the dark matter halo treated in this project. While halo density profiles usually follow the typical shape of an NFW profile, which in general shows two regions of different slopes, this is poorly visible in this case, due to the transition being at a small radius for this profile. The innermost region show some noise, caused by few counts in those bins.

weight proportional to the number of particles in each bin, giving less weight to the innermost bins, as those in this halo on the whole contain fewer particles than outer bins. The final density profile with the fitted curve using the least squares fitting method is shown in figure 3.2.

Intuitively, the fit seems very good for the majority of the outer bins but deviates a lot from innermost data points. This would make sense, as the latter have largest error bars and hence only loosely constrain the fit. This is because with increasing size of the error bar, the number of possible ways that the fitted curve can pass through increases as well. However, as was found in a next step described in section 3.2.1, the goodness of fit was in fact better for the inner bins than for the outer ones.

With regard to the error bars in the innermost region of the density profile, it can be seen that they are considerably larger than the rest, the ones further out being so small that they are not even visible in the figure. The large ones for the innermost data points are almost as large as the absolute magnitude of the data points themselves, which is merely a plotting artefact from plotting in logarithmic space. Specifically, each data point is given by $\rho \pm \sigma$. Subtraction of an error bar of the size of almost the data point itself gives a value close to zero, which explains why the error bars go as far down in the profile.
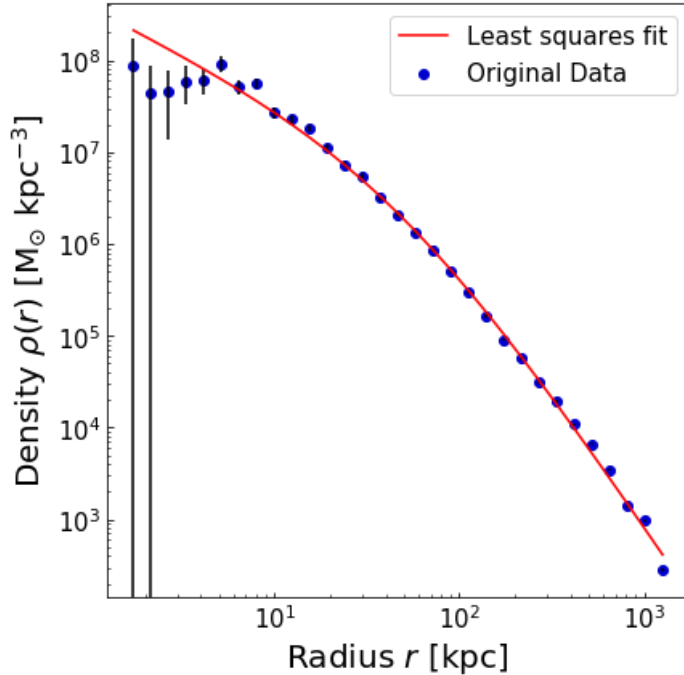
Figure 3.2: The figure shows the weighted least squares fit of the obtained halo density profile. Assuming Poisson error counts, an error bar was given to each data point, where the innermost bins have the largest error bars, due to having the fewest particles. Vertical artefacts in the innermost bins are caused by our assumption of symmetric Gaussian error bars, turning the density negative on a logarithmic scale. Error bars in the outer bins are smaller than the size of the points and thus cannot be seen.

3.2.1 Calculating the goodness of fit for individual data points

There are situations in which the interest does not lie in χ^2 itself, but instead, the individual values of $((y_i - f_i(x))/\sigma_i)^2$ in equation (2.5) are calculated, without taking the sum thereof. This gives the goodness of fit of each data point and can hence be used for the investigation of whether any data points form outliers.

After having fit the data with least squares, the goodness of fit was calculated for each individual data point of the halo density profile and represented with different colours in figure 3.3. Each colour corresponds to a value, indicated by the colour bar on the right-hand side in the figure, where a smaller value signifies a better fit. This was done in order to investigate whether it was indeed the innermost points that had the worst fit, as had initially been suspected.

The figure shows that the goodness of fit in fact gets worse with increasing radius. It was previously noted that the error bars are large for the innermost data points, while at larger radii, they are too small to be visible. This means that the fitted curve is heavily constrained by the outer data points and only loosely constrained by the inner ones. Since $((y_i - f_i(x))/\sigma_i)^2$ in the expression for χ^2 represents the distance between a data point and the fit, weighted by the size of the error bar in that point, it is possible that the

fit visually seems good, despite this not being true due to having a very small error bar, which increases the value of $((y_i - f_i(x))/\sigma_i)^2$. Conversely, a fit might look bad, as the curve might be far from the data points, but still be decently good, due to a large error bar at that point. These two cases are exactly what is seen in figure 3.3 for the outermost and innermost data points respectively.
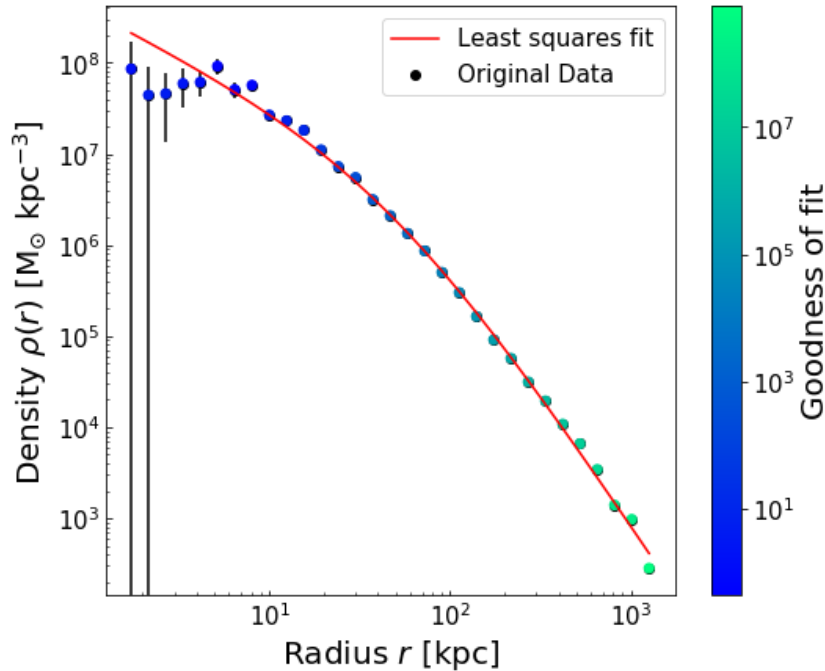

Figure 3.3: By calculating the goodness of fit for each data point, which in the figure is illustrated by different colours, it can be seen that the fit gets worse with increasing radius. This is caused by the uncertainties being underestimated in these points and hence, impose a too harsh restriction on the fit.

There exist several conclusions and interpretations of these observations. First of all, seeing that there is a heavy constraint on the fitted curve for the outermost bins and that the fit is bad for those points, the interpretation was made that the error bars must have been underestimated. This, along with the alternative interpretation that the used fitting method could have been wrong, is briefly discussed in chapter 4.1. Going with the first interpretation, it can be said that while Poisson error bars are a decent description of the uncertainties, there seem to be additional sources of error, which were not taken into account. Modelling such additional uncertainties, beyond the contribution of Poisson noise, is however outside the scope of this project. This assumption is kept throughout the rest of this report, but future work could include taking additional sources of error into account, as will be discussed in chapter 4.2. Apart from that, it can be argued that it is not fully accurate to describe the uncertainties by Poisson counts, as this assumes each bin to be independent of the others. Realistic halos are in fact constituted of substructures, making it more likely for a bin to have a high density if that was the case for the neighbouring

one. A further assumption that was made in this respect, was that the number of particles in each bin was large enough to assume the Poisson distribution to be symmetric. This is however not fully accurate and contributes to the observed result.

A further explanation for the observed goodness of fit is linked to the model itself. While the NFW profile is known to fit halo density profiles well, the fit was not expected to be perfect, as realistic halos have large amounts of substructures. Additionally, the expression for the NFW profile diverges to infinity as the halo radius approaches zero, which leads to the expectation of a bad fit at small radii. However, this problem is resolved by the fact that there are limitations to the resolution of the simulation, which the data is based on. For large radii on the other hand, the expression approaches zero, whereas realistic halos must be truncated by defining an outer edge, commonly by using the halo's virial radius. Hence, the model also breaks down at the outermost radii of the density profile and the fit is not expected to be good there either. The latter could be observed for the outermost data points in this density profile in figure 3.3, while at the innermost radii, the large error bars compensated the bad fit, as previously mentioned.

3.2.2 Likelihood from least squares for a fixed parameter space

In a next step, the likelihood was calculated for multiple sets of parameter value combinations, given by a 300 x 300 parameter space that was constructed from a vast range of values. Specifically, the x-axis represented values for r_s, ranging between $1 \cdot 10^1 - 2 \cdot 10^2$ kpc, while the values for ρ_0 on the y-axis ranged between $5 \cdot 10^5 - 7 \cdot 10^7$ M$_\odot$ kpc^{-3}. Equation (2.7) was then applied to the value of χ^2 for the set of parameter values in each grid point, in order to determine those that give the best fit of the model to the data. Not only does this method enable gaining the full probability distribution of the parameters, but also it enables the verification of whether or not the probability distributions of the likelihoods are single-peaked. This could not be investigated from the least squares method alone, as the latter merely gives a single value for each parameter. These probability distributions representing the likelihood of the parameters could then be used when applying Bayesian inference at a later point.

The previous application of the least squares fitting method gave the set of parameters that minimized χ^2. The knowledge about these values was used for defining the ranges of the parameter space, making sure that they were located near the centre of the grid. While the initial sampling grid was defined very wide, a more restricted range was used for the contour plot, once the sets of parameters giving the maximum likelihood had been found. The likelihood, which was calculated for each grid point, was then plotted in a contour plot shown in figure 3.4. It was defined in a way that similar values for the likelihood are represented as similar colours, where a larger value means a larger likelihood and hence, a better fit between model and data.

The contour itself shows a highly eccentric and skewed ellipse of multiple colours, which visualizes the likelihood for different sets of parameters. As it deals with two parameters, represented by either axis, the problem is said to be two-dimensional, the contour plot being the joint, two-dimensional probability distribution for the two parameters. The cross

represents the set of parameter values from least squares that gives the best fit between model and data. As expected, it coincides with the central region of the ellipse, which has the largest value for the likelihood. The observation can be made that there are various parameter value combinations sharing the same likelihood. Additionally, each colour in
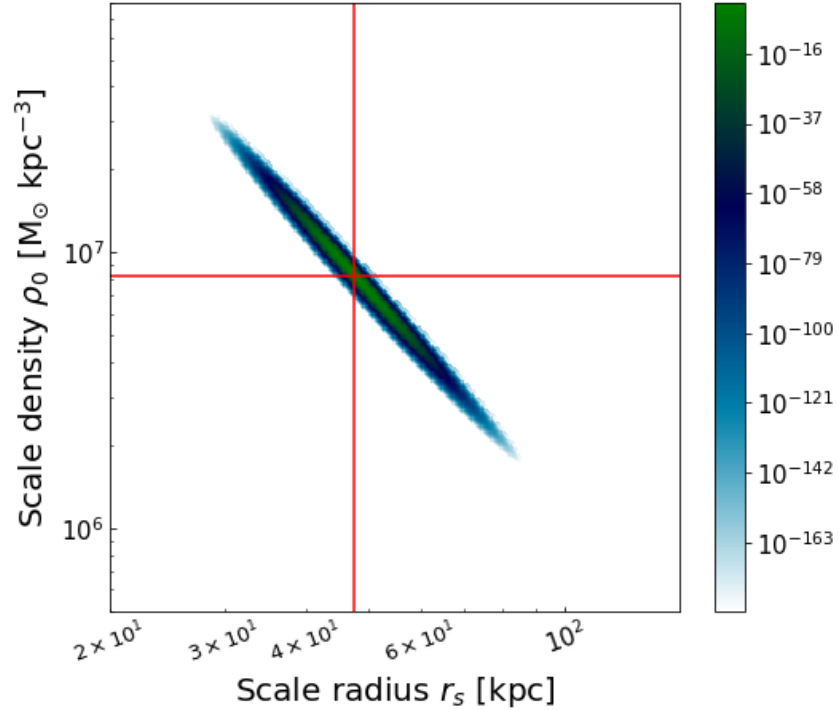

Figure 3.4: The figure shows the likelihood of the model being a good fit to the data for each point in the parameter space. The likelihood is indicated by different colours, increasing towards the centre of the ellipse. The cross represents the most likely parameters according to the least squares fit.

the ellipse has a significant width, which is linked to an uncertainty. The parameter values that give rise to a specific colour in the contour are a good fit to a specific value within a certain confidence level, which gradually increases with distance from the centre of the ellipse. It is important to keep in mind that the uncertainties have not been explicitly defined to fit the confidence levels that are common for Gaussians, but rather that they merely increase outwards from the centre of the ellipse. In fact, the contour has been highly extended in terms of contrast, meaning that the contour extends far outside the Gaussian confidence levels. From the observations concerning regions in the contour sharing the same colour, the conclusion can be drawn that there is a degeneracy of parameters, as parameter combinations leading to the same value for the likelihood have equal probability to give the model a good fit to the data. Another way of expressing it, is that several combinations of values for the parameters give an equally good fit at a certain σ level.

A further point of interest is the contour's shape. Had the contour been circular, the two parameters would have been uncorrelated. In the present likelihood contour however, the contour is both highly eccentric and skewed, indicating a strong correlation between the

two. The elliptical shape further enhances the degeneracy of the parameters, as opposed to the case of a circular contour, where the parameters would not have been as degenerate and would have had almost equal probabilities and similar uncertainties. The fact that the ellipse's semi-major axis is considerably larger than its semi-minor axis makes the uncertainties considerably larger in the direction of the former axis. Along the semi-minor axis on the other hand, the regions of equal colour and thus, the uncertainties are much narrower and therefore, the parameters are much more constrained in that direction. Hence, there is a much stronger degeneracy along the major axis, since it includes sets of parameters that differ largely in their values, but still make the model fit the data equally well. On the whole, a larger compression of the ellipse implies a stronger degeneracy.

3.3   Bayesian inference

The second approach used for finding the most likely values of the parameters and their respective uncertainties, which is based on the first one, was made by applying Bayesian inference. This involved multiplying the previously found likelihood by prior knowledge about the parameters, in order to obtain a posterior distribution. Initially, a flat prior was used, before applying a physical prior based on knowledge about the relation between halo masses and concentrations instead.

3.3.1   Inference using a flat prior

Equal probability was given to all parameters in the previously defined parameter space, which was found from equation (2.10), and zero probability was given to the values outside of it. Figure 3.5 shows this prior in the leftmost panel. As expected, the parameter space is unicoloured, meaning that each grid point has equal probability. The colours in the contours correspond to the values given by the colour bar in figure 3.4, which holds true in all figures unless otherwise stated. In the panels in the center and to the right of the same figure, the likelihood and the posterior are shown. A comparison between the two shows that they are identical, as was expected from the use of a flat prior. The reason for this, is that the prior did not contribute with any information about the parameters and hence, had the effect of multiplying the likelihood by 1. A flat, wide prior is equivalent to multiplying the likelihood by 1 and hence, producing a maximum posterior that is not shifted compared to the maximum likelihood. This shows that it makes sense to induce prior knowledge about the parameters when applying Bayesian inference, if that is available, as was done in the next step. A flat prior should only be used if there is a lot of initial ignorance about the parameters.
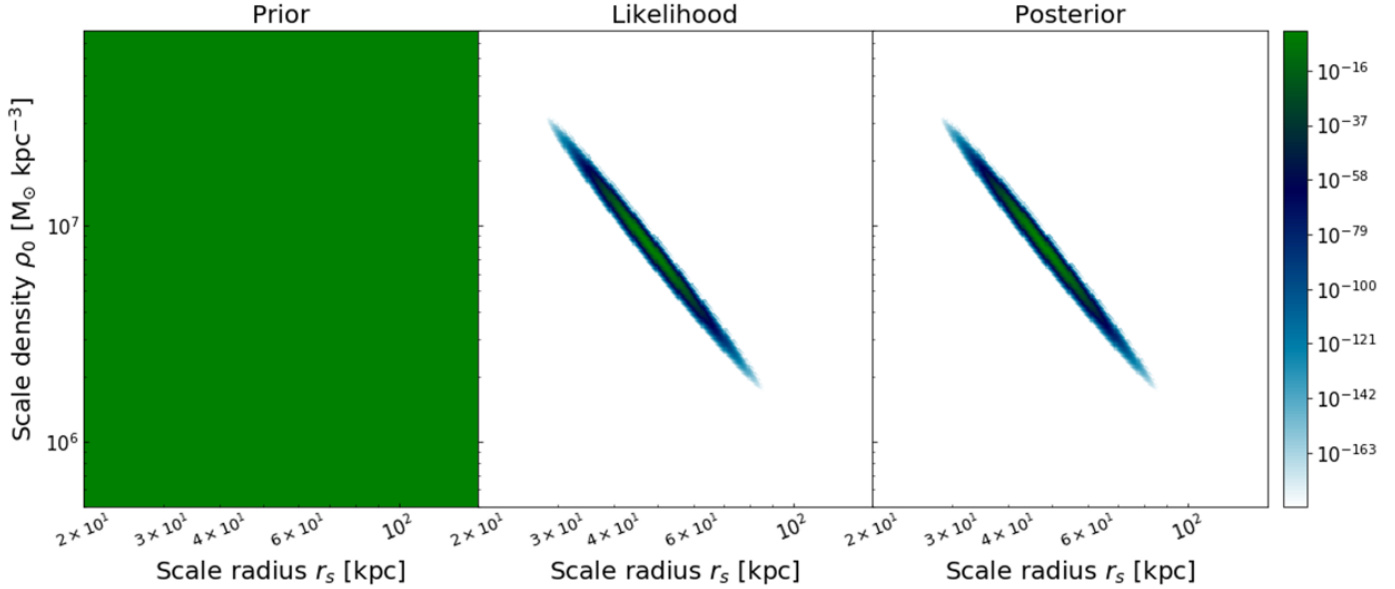
Figure 3.5: The different panels in the figure show the prior, likelihood and the resulting posterior from the application of Bayesian inference. A flat prior was used, which explains why the likelihood and the posterior are identical. A flat prior does not contribute to the final result, as it is equivalent to multiplying the likelihood by 1.

3.3.2   Inference using a physical prior

As could be seen in the previous step, the application of a flat prior resulted in the likelihood remaining unchanged. By using a physical prior instead, previous knowledge about the parameters can be embedded, which sets constraints on the possible parameter values. In this project, this was deduced from the knowledge about the mass-concentration relation of dark matter halos (Dutton & Macciò 2014), which underlies the fact that halos can only have a limited range of values for the concentration, given their mass. The following sections describe the procedure of finding such typical values for the parameters and their uncertainties for an average halo of equal mass as the one treated in this project.

First of all, the virial mass and virial radius were found for this halo. Based on the fact that the latter is defined as the radius at which the density of the halo equals 200 times the critical density of the Universe, while the virial mass is given by the mass enclosed by this radius, these values were found to be $M_{200} = 1.38 \cdot 10^{13}$ $M_\odot$ and $r_{200} = 508.522$ kpc. The afore-mentioned relation between halo mass and concentration was taken from figure 7 in Dutton & Macciò (2014). This relation could be used, as it applies to halos with similar cosmological parameters as the ones in the simulation from which the density profile in this project was taken. The fit in the previously mentioned figure is based on equations 7 and 8 in Dutton & Macciò (2014), which give an approximation of the median value of a halo's concentration, given its mass, followed by an expression for the best fit. Inserting the previously found virial mass, the median concentration was found to be 6.416. A comparison of this with the concentration of the halo of this project by using equation (1.3) showed

that the latter is considerably more concentrated than the average halo. This deviation between the two values is however no reason for concern, as there exists a scatter for the concentration. The obtained value from this mass-concentration relation merely describes a statistical average, meaning that it might not be true for every halo, including the one treated in this project, but rather, it represents the concentration of an average halo with that mass.

After finding that $c = 6.416$ is the mean of the concentration of an average halo with mass equal to the virial mass of the halo treated here, the next step was to find the uncertainty thereof. The afore-mentioned scatter in halo concentration about the median is given by 0.11 dex for a given mass (Dutton & Macciò 2014). The translation thereof into the expected range of the concentration for a halo of mass $M_{200} = 1.38 \cdot 10^{13}$ $M_\odot$ resulted in the values given in (3.1) and (3.2), where $median_c$ is the median of the concentration for a halo of that mass.

$$c + 1\sigma = 10^{\log(median_c)+0.11} = 8.26 \tag{3.1}$$

$$c - 1\sigma = 10^{\log(median_c)-0.11} = 4.98 \tag{3.2}$$

The use of dex is given by the fact that equation 7 in Dutton & Macciò (2014) uses the logarithm of the concentration and mass. The reason for this is because it compresses the values, which is necessary, as they range over several orders of magnitude. The obtained values in equations (3.1) and (3.2) mean that the average halo with mass $M_{200} = 1.38 \cdot 10^{13}$ $M_\odot$ has an average concentration of 6.416, where 68% of its mass correspond to a concentration between 4.98 and 8.26. This is based on the assumption that the probability distribution can be described by a Gaussian function, and hence, the range given by $\mu \pm \sigma$ constitutes 68% of the area under the curve (Sivia & Skilling 2006).

The final step of retrieving values for the parameters and their uncertainties, was to derive them based on the found values for the mass and concentration. The parameter r_s was found by inserting the obtained values for the concentration and virial radius in equation (1.3). For obtaining the value for ρ_0 on the other hand, the fact was used that the NFW profile can be described by two different expressions. Up until now, the expression for the density with respect to radius, as stated in equation (1.1), has been used. Alternatively however, equation (3.3) can be used instead to fully describe the density profile. In this equation, the density profile is integrated up to a certain radius and describes the mass enclosed by a sphere of that radius.

$$M = \int_0^{r_{max}} 4\pi r^2 \rho(r) dr = 4\pi \rho_0 r_s^3 \left[\ln\left(\frac{r_s + r_{max}}{r_s} - \frac{r_{max}}{r_s + r_{max}}\right) \right] \tag{3.3}$$

In order to obtain the virial mass, r_{max} can be replaced by the virial radius. Additionally, the equation can be rewritten in terms of concentration according to equation (1.3), which

leads to the expression in equation (3.4). This is an adapted version of equation 7.139 in Houjun Mo & White (2010).

$$M_{200} = \int_0^{r_{200}} 4\pi r^2 \rho(r) dr = 4\pi \rho_0 r_s^3 \left[\ln\left(1 + cx\right) - \frac{cx}{1 + cx} \right],$$ (3.4)

where $x \equiv r/r_{max}$. By setting $x = 1$, the expression gives the enclosed mass (Houjun Mo & White 2010). Inserting the previously found values for r_{200}, M_{200}, r_s and c gives the value of ρ_0.

These parameter values, which hold for an average halo having the virial mass of the halo treated in this project, were through application of the above procedures found to be given by $r_s = 80$ kpc $\pm$ 20 kpc and $\log(\rho_0) = 6.3$ M$_\odot$ kpc$^{-3} \pm 0.25$ M$_\odot$ kpc^{-3}. As can be seen, the values were found for a Gaussian probability distribution in linear space for r_s, while they were found for a Gaussian probability distribution in logarithmic space for ρ_0. The reason for doing so, is because the former parameter ranges over several orders of magnitude and hence, a logarithmic spacing compresses this. A consequence of doing so, was that it had to be translated back into linear space again, before applying it to the prior grid. This translation is shown in equation (3.5). Since the uncertainties were symmetric in logarithmic space, they no longer are in linear space.

$$P(\rho_0) = 10^{6.3\,\pm\,0.25} \text{ M}_\odot \text{kpc}^{-3} = 2 \cdot 10^6\ + 1.5 \cdot 10^6 \text{ M}_\odot \text{kpc}^{-3}$$
$$- 9.0 \cdot 10^5 \text{ M}_\odot \text{kpc}^{-3}$$ (3.5)
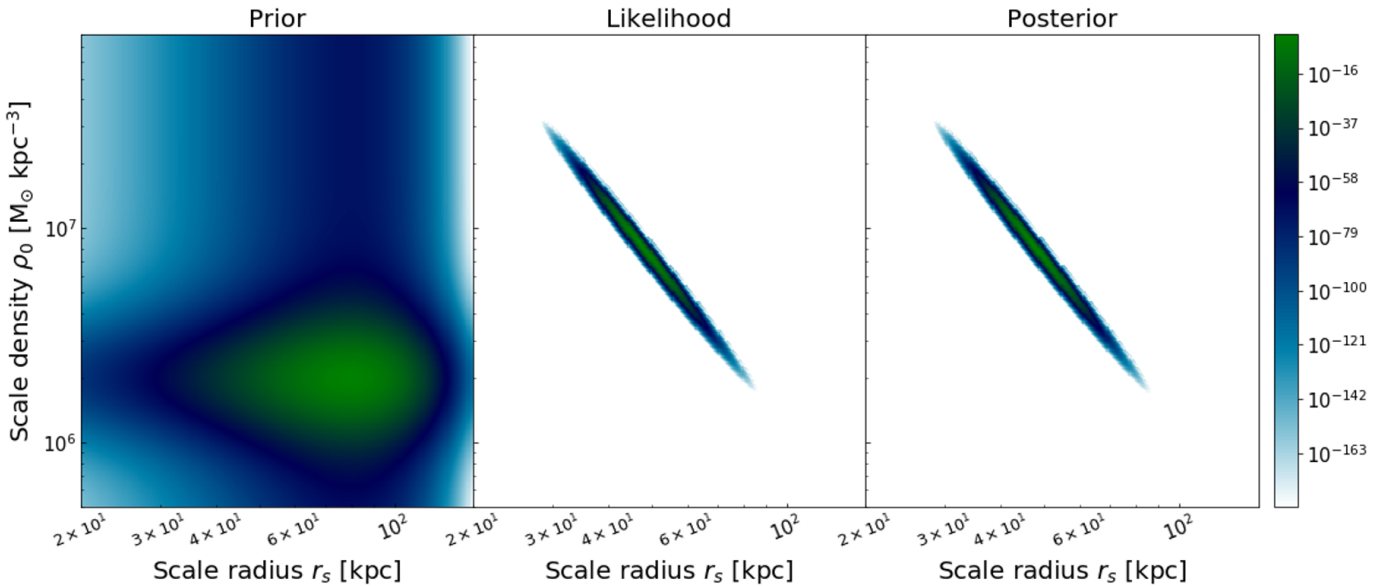

Figure 3.6: A physical prior was found from a known mass-concentration correlation, which sets restrictions on the possible parameter values, as indicated by the green region in the leftmost panel. Nevertheless, no difference can be observed between likelihood and posterior, arising from a strong underestimation of the uncertainties. The contour is defined as in figure 3.4

The values for the parameters were rounded, as the prior should be only broadly informative. They were inserted in Gaussian functions, which were multiplied by one another. For each grid point in the parameter space, the resulting values were then plotted, as shown in the leftmost panel of figure 3.6. It shows different regions, denoted by different colours, where largest probability is given to the central region illustrated in green. The central and the rightmost panel in the figure show the likelihood and the posterior respectively. Despite having used a physical prior, there is no visible difference between the contours of the likelihood and the posterior. This is surprising, as the most likely values for ρ_0, given by the prior, are considerably smaller than those given by the likelihood, while for r_s, the prior shows larger values. Hence, while the prior was chosen to be fairly broad, as it should make little impact on the posterior, an effect of the prior on the posterior would have been expected, since the central regions of the contours for the prior and the likelihood differ so strongly. The reason for the absence of a shift between the likelihood and the posterior distribution was found to be a consequence of numerical inaccuracy in the following way. In regions where the prior is non-zero, the likelihood is so small, that it is taken to be zero, due to limitations to the numerical accuracy. Since the posterior is found by multiplication of the prior with the likelihood, this becomes zero in those regions, leaving the posterior to show the same contour as the likelihood. This shows that despite using a physically motivated prior, its mean and sigma are so far away from the values that would be allowed by the likelihood, that it becomes uninformative. The only regions in which this prior would give any information about the parameter values are ones where the model has zero probability of fitting the data. From this observation, the conclusion was drawn that the uncertainties in the likelihood had been underestimated. In this project, the uncertainties of the data were assumed to be dominated by Poisson noise. However, there must be an additional source of error, which is considerably bigger than the Poisson noise, that was not taken into account. If the uncertainties had not been underestimated, the error bars in the likelihood contour would have been bigger, meaning that the Gaussian confidence level would have extended further out, giving a larger likelihood to parameter values that are currently within the white space of the likelihood contour and hence, currently have zero probability of fitting the data.

As stated earlier, the contours in the figures up to this point were not defined explicitly and were found to range over as much as 36σ, which is considerably outside the Gaussian confidence levels. Specifically, the outermost colour in the contours in for example figure 3.6, excluding the white regions, corresponds to approximately 30σ away from the mean of the distribution. Taking this into consideration when comparing the central regions of the contours of the prior and the likelihood in the afore-mentioned figure gives that the most likely parameters given by the physical prior are many σ away from the best-fit parameters given by the likelihood contour. This further underlines the fact that the uncertainties of the data and hence of the likelihood were underestimated.

Finally, the marginalized distribution of the posterior for each of the parameters was found. These are shown in the top and bottom right panel in figure 3.7 respectively, along with the two-dimensional posterior distribution in the bottom left panel. Since these one-dimensional distributions were assumed to follow Gaussian probability distribution

functions, but the contours were defined over so many σ levels, the resulting distributions were misleadingly narrow, suggesting small uncertainties in the posterior. In order to get more reasonable Gaussian distributions, but avoid having to calculate the σ contours explicitly, the contours of the two-dimensional posterior distribution was redefined in such a way, that the contour is restricted to 5σ. This means that the outermost colour excluding the white regions corresponds to a 5σ level. Additionally, the x-axis showing values for r_s was set to linear spacing, as were the values of the colours in the contour. As can be seen in figure 3.7, this enabled the Gaussian distributions to run over the full width of the contour. Moreover, the best-fit parameters that were found from least squares were indicated as a cross in the two-dimensional posterior distribution and as lines in the two marginalized distributions of the posterior. As was noted previously, there is no shift to be seen between the likelihood and the posterior, which can be taken from the fact that the best-fit parameter from the likelihood coincides with the mean of the posterior distributions.
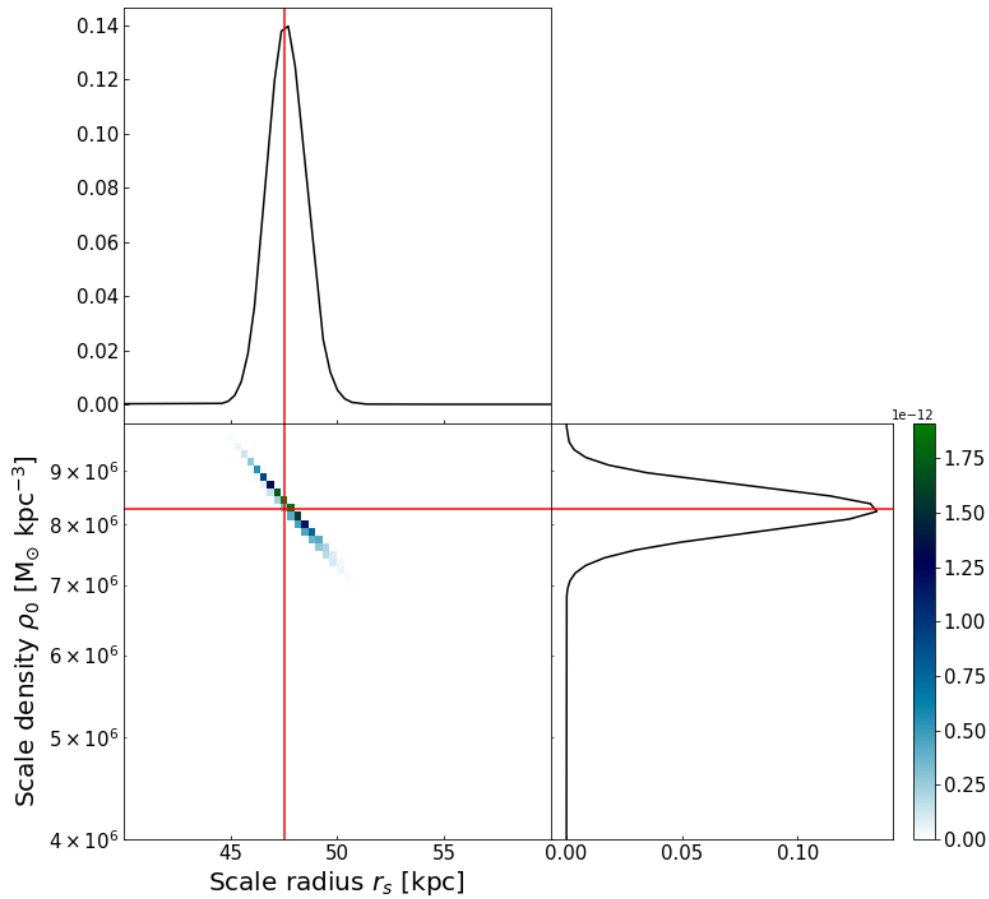

Figure 3.7: The figure shows the contour of the posterior distribution in the bottom left panel, along with the marginalized, one-dimensional distributions of each parameter respectively in the remaining panels. The contour was restricted to 5σ levels, in order to obtain reasonable Gaussian distributions.

While the grids constituting the parameter space had initially been made using a 100 x 100 parameter grid, the grid was increased to 300 x 300 upon noticing that the contour of the posterior distribution in figure 3.6 else becomes highly pixelated and hence, badly resolved when the contour is restricted to 5σ levels only. This is a strong indication of the sampling method being highly inefficient. The inefficiency arises from the fact that the sampling was done using fixed ranges for the axes of the grid and that the majority of values retrieved with this procedure are insignificant, creating the white regions in the contour plot of for example figure 3.7. Despite only a small region contributing to the best-fit parameters, it was necessary to start with a broad sampling space, because this could only be known once the sampling was completed. The region that gives rise to these parameters, which is the region of highest interest for this problem, is however only poorly resolved. More efficient samplers are briefly discussed in chapter 4.2, which is an improvement left to future work.

Chapter 4

Discussion and Conclusion

4.1 Comparison to literature method

As was stated in chapter 1, the procedure of fitting the NFW model to halo density profiles by minimizing ψ^2 in equation (1.2), as was previously done in the literature (Dutton & Macciò 2014), did not include data uncertainties. Therefore, this project dealt with determining the set of parameter values that gives the best fit between this model and a halo density profile, by actually taking these uncertainties into account.

In order to have a direct comparison between the literature method and the method used in this project and hence, to be able to see the effect of error bars of the data points on the fit, the method was applied to the density profile treated in this project. As can be seen from equation (1.2), this involved subtracting the logarithm of the data by the logarithm of the NFW function applied to the data and in that way find the most likely parameter values. This procedure is analogous to least squares, with the only exception that it is executed for logarithmic spacing and that all data points are weighted equally. Using a logarithmic spacing compresses the values, making any variations more subtle and the fit more well-behaved. The obtained values for the fitted curve were then converted back into linear space, in which they were plotted. The resulting curve fit along with the least squares fit used in this project are shown in the left panel in figure 4.1 for the sake of comparison. The figure shows that the fits are overlapping in outer regions of the halo, while they deviate slightly in the inner regions. The deviation arises from the fact that the literature method does not include uncertainties, meaning that each bin is weighted equally. Conversely, when applying least squares in this project, each data point was weighted by a Poisson error bar based on the number of counts, which on the whole gave outer bins more weight, as those in most cases included more particles. Therefore, a first conclusion is that the literature method gives too much weight to the inner bins.

When applying different fitting methods, the aim is to fit the data as well as possible, without actually knowing what the true fit is. When using the weighted least squares method, error bars were given to each point. If their size is correctly estimated, the true fit is expected to lie within these error bars. Despite the literature method not weighting
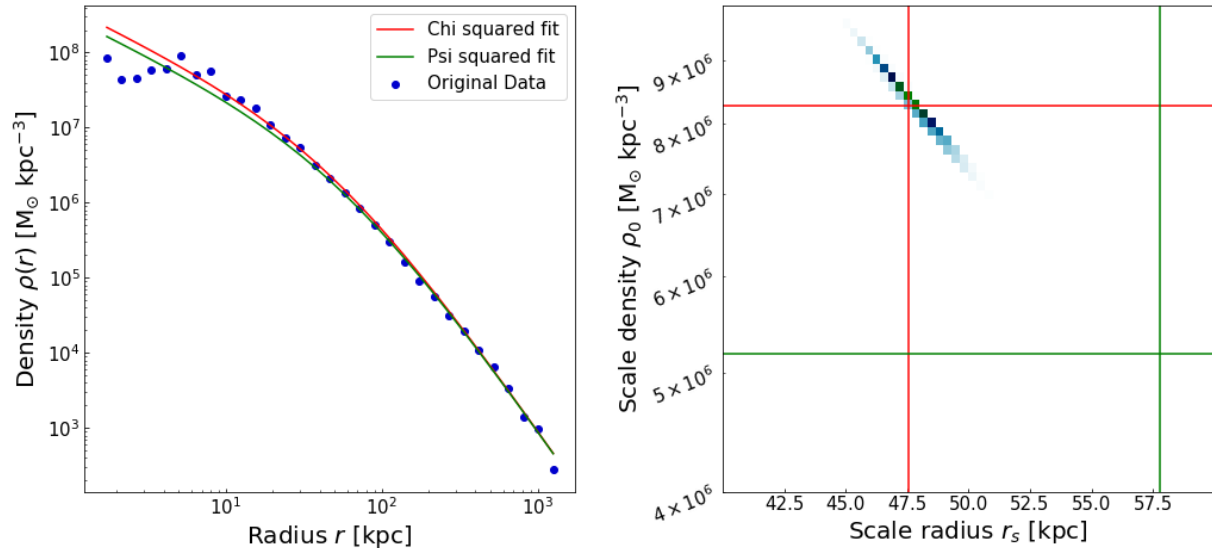
Figure 4.1: In the left panel, the weighted least squares fit used in this project (denoted as Chi squared fit) is compared to an unweighted analogue used in literature (denoted as Psi squared fit). In the right panel, the optimal parameters from least squares (red) and those from the literature method (green) are indicated as crosses in the likelihood contour, indicating a strong underestimation of the confidence intervals.

the bins when making the curve fit, it was expected to give a similar fit as the one from least squares. From the left panel in figure 4.1 it can be seen that the two are indeed similar, as was addressed in the previous section. Comparing the best-fit parameters gained from the two methods for the halo treated in this project, gives the following. The previously found parameter values from least squares, where χ^2 was minimized, are given by $r_{s_{\chi^2}} = 4.753 \cdot 10^1$ kpc and $\rho_{0_{\chi^2}} = 8.266 \cdot 10^6$ M$_\odot$ kpc^{-3}, as opposed to the ones from ψ^2, which are given by $r_{s_{\psi^2}} = 5.779 \cdot 10^1$ kpc and $\rho_{0_{\psi^2}} = 5.184 \cdot 10^6$ M$_\odot$ kpc^{-3}. In the right panel in figure 4.1, these best-fit parameters for each of the two cases are indicated by crosses in the likelihood grid taken from the bottom left panel in figure 3.7. The contour shows the likelihood on a 5σ level, as was found using the likelihood in each point, based on the values from χ^2 and the assumption of Poisson noise as the dominant source of uncertainties. However, while the observation was made that the two fits looks similar in the left panel, the best-fit parameters obtained from the literature method are found to be far outside the confidence levels given by the data uncertainties in the right panel. In fact, the two set of parameters are well outside the 5σ level of the likelihood contour and incompatible even on a 10σ level. This is surprising, as the statistical error from the Poisson noise was expected to be large enough to encompass the systematic error. From this, the conclusion was drawn that while the two fits look visually compatible in the left panel, they are in fact not in a statistical sense. This means that there are additional sources of error that were not taken into account, as was found already in section 3.3.2, which lead to an underestimation of the confidence interval. From the above observations, these additional sources of error are in fact larger than the Poisson noise, which had initially been assumed to be the leading

uncertainty. This is because the systematic errors were found to be a much stronger source of uncertainty than the statistical Poisson noise, as the two different methods give strongly incompatible results. Had the uncertainties not been underestimated, the contour would have been inflated to larger regions in the chosen parameter space, including the physical prior used in section 3.3.2 as well as the best-fit parameters from the literature method.

4.2 Future research

The following sections describe a set of improvements to the current procedures of this project, followed by future work that could be done, using the current work as a basis. With regard to improvements, a first point about the error bars of the data points can be made. When assigning uncertainties to the data points, the assumption was made that the errors follow a Poisson distribution and that the number of counts within one bin was big enough for the assumption to be made that these Poisson distributions were symmetrical. As a matter of fact, this was not a fully accurate description of the uncertainties, as was predominantly visible in figure 3.3, showing the goodness of fit for individual data points. First of all, it was a simplification to assume the counts to follow a Poisson distribution, because the latter assumes each bin to be independent of the others. Instead, halos are constituted of substructures, meaning that neighbouring bins are likely to have similar values for their density. Moreover, the number of particles in the innermost bins in particular were in fact not sufficient for the assumption to be made, that the Poisson probability distribution is symmetric. Instead of making these two assumptions, the real shape of the Poisson distribution could be implemented. Furthermore, the assumption was made throughout this project that a halo is spherically symmetric, despite this not generally being the case. As a consequence, additional sources of error are to be expected apart from the Poisson counts of each bin, which implies a correlation between bins. This could be embedded in the likelihood through the use of a covariance matrix. Finally, as was found from the use of a physical prior, but also when comparing the weighted least squares method with the literature method, the Poisson noise was in fact not the dominant source of error. Hence, an important task for future work would also be to investigate what the main cause of the uncertainties actually is and to iterate the whole procedure by taking these further sources of error into account.

With respect to the likelihood, there exist alternative ways of finding the set of parameter values that give the best fit between model and data. In this project, a parameter space consisting of a square grid was chosen as the sampling. This was however found to be highly inefficient, as the majority of the grid points had a vanishing likelihood, which could only be known once the sampling was completed. Hence, a lot of time was spent calculating the likelihood of grid points that clearly did not contribute to the maximum likelihood. An alternative method would be to use an algorithm as for example the Markov chain Monte Carlo (MCMC) algorithm, in order to increase the efficiency of the sampling. Despite this method being more efficient, it is important to underline that it does not provide a greater

precision than the one used in this project.

The next step to what was done in this project would be to first of all, apply the whole procedure of retrieving halo density profiles and finding best-fit parameter values for the remaining halos of the halo population in the simulation that was used. On the one hand, this would give a more complete picture of the typical parameter values and their uncertainties for dark matter halos of various sizes. On the other hand, the found values for r_s could then be used to find the halos' concentrations from equation 1.3, which in turn could be plotted with respect to the halo masses. This would help to gain a better understanding of the response of the mass-concentration relationship to the propagation of the inferred uncertainties to halo concentrations. As mentioned in chapter 1, the mass-concentration relationship has previously been recovered for a halo population of similar characteristics (Ludlow et al. 2016). The crucial difference was however, that the uncertainties of the data points and hence, of the best-fit parameters were neglected, giving only point estimates for the concentrations. The interest when including uncertainties lies in the investigation of whether or not the incorporation of error bars to the concentrations has an effect on the trend between halo masses and concentrations.

Acknowledgements

Bibliography

Aksu, A. (2018), *Decision Theory and Bayesian Analysis.*

Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C. & Zaritsky, D. (2006), 'A Direct Empirical Proof of the Existence of Dark Matter', **648**(2), L109–L113.

Davis, T. M. & Parkinson, D. (2016), 'Characterising Dark Energy through supernovae', *arXiv e-prints* p. arXiv:1610.09452.

Dutton, A. A. & Macciò, A. V. (2014), 'Cold dark matter haloes in the Planck era: evolution of structural parameters for Einasto and NFW profiles', **441**(4), 3359–3374.

Fageot, J., Uhlmann, V. & Unser, M. (2017), 'Gaussian and Sparse Processes Are Limits of Generalized Poisson Processes', *arXiv e-prints* p. arXiv:1702.05003.

Haslwanter, T. (2013), 'Residuals for linear regression fit'.
URL: *https://commons.wikimedia.org/wiki/File:ResidualsforLinearRegressionFit.png*

Houjun Mo, F. v. d. B. & White, S. (2010), *Galaxy Formation and Evolution*, Cambridge University Press.

Jing, Y. P. (2000), 'The Density Profile of Equilibrium and Nonequilibrium Dark Matter Halos', **535**(1), 30–36.

Leistedt, B. (2017), 'Fitting a line to data - a quick tutorial'.
URL: *https://ixkael.github.io/fitting-a-line-to-data-a-quick-tutorial/*

Ludlow, A. D., Bose, S., Angulo, R. E., Wang, L., Hellwing, W. A., Navarro, J. F., Cole, S. & Frenk, C. S. (2016), 'The mass-concentration-redshift relation of cold and warm dark matter haloes', **460**(2), 1214–1232.

Navarro, J. F. (1995), 'The Structure of CDM Halos', *arXiv e-prints* pp. astro–ph/9511016.

Navarro, J. F., Frenk, C. S. & White, S. D. M. (1996), 'The Structure of Cold Dark Matter Halos', **462**, 563.

Navarro, J. F., Frenk, C. S. & White, S. D. M. (1997), 'A Universal Density Profile from Hierarchical Clustering', **490**(2), 493–508.

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J. P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J. F., Carron, J., Challinor, A., Chiang, H. C., Chluba, J., Colombo, L. P. L., Combet, C., Contreras, D., Crill, B. P., Cuttaia, F., de Bernardis, P., de Zotti, G., Delabrouille, J., Delouis, J. M., Di Valentino, E., Diego, J. M., Doré, O., Douspis, M., Ducout, A., Dupac, X., Dusini, S., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen, H. K., Fantaye, Y., Farhang, M., Fergusson, J., Fernandez-Cobos, R., Finelli, F., Forastieri, F., Frailis, M., Fraisse, A. A., Franceschi, E., Frolov, A., Galeotta, S., Galli, S., Ganga, K., Génova-Santos, R. T., Gerbino, M., Ghosh, T., González-Nuevo, J., Górski, K. M., Gratton, S., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Handley, W., Hansen, F. K., Herranz, D., Hildebrandt, S. R., Hivon, E., Huang, Z., Jaffe, A. H., Jones, W. C., Karakci, A., Keihänen, E., Keskitalo, R., Kiiveri, K., Kim, J., Kisner, T. S., Knox, L., Krachmalnicoff, N., Kunz, M., Kurki-Suonio, H., Lagache, G., Lamarre, J. M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Le Jeune, M., Lemos, P., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Lilley, M., Lindholm, V., López-Caniego, M., Lubin, P. M., Ma, Y. Z., Macías-Pérez, J. F., Maggio, G., Maino, D., Mandolesi, N., Mangilli, A., Marcos-Caballero, A., Maris, M., Martin, P. G., Martinelli, M., Martínez-González, E., Matarrese, S., Mauri, N., McEwen, J. D., Meinhold, P. R., Melchiorri, A., Mennella, A., Migliaccio, M., Millea, M., Mitra, S., Miville-Deschênes, M. A., Molinari, D., Montier, L., Morgante, G., Moss, A., Natoli, P., Nørgaard-Nielsen, H. U., Pagano, L., Paoletti, D., Partridge, B., Patanchon, G., Peiris, H. V., Perrotta, F., Pettorino, V., Piacentini, F., Polastri, L., Polenta, G., Puget, J. L., Rachen, J. P., Reinecke, M., Remazeilles, M., Renzi, A., Rocha, G., Rosset, C., Roudier, G., Rubiño-Martín, J. A., Ruiz-Granados, B., Salvati, L., Sandri, M., Savelainen, M., Scott, D., Shellard, E. P. S., Sirignano, C., Sirri, G., Spencer, L. D., Sunyaev, R., Suur-Uski, A. S., Tauber, J. A., Tavagnacco, D., Tenti, M., Toffolatti, L., Tomasi, M., Trombetti, T., Valenziano, L., Valiviita, J., Van Tent, B., Vibert, L., Vielva, P., Villa, F., Vittorio, N., Wand elt, B. D., Wehus, I. K., White, M., White, S. D. M., Zacchei, A. & Zonca, A. (2018), 'Planck 2018 results. VI. Cosmological parameters', *arXiv e-prints* p. arXiv:1807.06209.

Pontzen, A., Roškar, R., Stinson, G. & Woods, R. (2013), 'pynbody: N-Body/SPH analysis for python'.

Rey, M. P., Pontzen, A. & Saintonge, A. (2019), 'Sensitivity of dark matter haloes to their accretion histories', **485**(2), 1906–1915.

Riemer-Sørensen, S. (2017), 'Statistics and model selection in cosmology'.
URL: *http://icg.port.ac.uk/ jschewts/cantata/L5/StatisticsNotes.pdf*

Sivia, D. S. & Skilling, J. (2006), *Data Analysis, A Bayesian Tutorial*, Oxford Science Publications.

Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E. & Hernquist, L. (2008), 'A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei', **391**(2), 481–506.

Springel, V. (2005), 'The cosmological simulation code GADGET-2', **364**(4), 1105–1134.

Wechsler, R. H. & Tinker, J. L. (2018), 'The Connection Between Galaxies and Their Dark Matter Halos', **56**, 435–487.

Appendix A

Derivation of the volume of a spherical shell

The following shows a derivation of the expression for the volume of a shell under the assumption that the approximation $\Delta r << r$ is valid. The volume of a sphere of radius r, which we call inner sphere, is given by

$$V_{inner} = \frac{4}{3}\pi r^3. \tag{A.1}$$

For a larger sphere of radius $r + \Delta r$, which we call outer sphere, the volume is given by the following, where merely the radius has been changed in the previous expression.

$$V_{outer} = \frac{4}{3}\pi (r + \Delta r)^3, \tag{A.2}$$

By subtraction of the outer sphere by the inner sphere, the following expression is obtained for the volume of a shell i.

$$V_i = \frac{4\pi}{3}\left[(r + \Delta r)^3 - r^3\right] \tag{A.3}$$

By application of the Taylor series expansion to first order, which is based on the initial assumption that $\Delta r << r$, the following expression is found.

$$\left(1 + \frac{\Delta r}{r}\right)^3 \simeq 1 + \frac{3\Delta r}{r} + \mathcal{O}\left(\frac{r\Delta r}{r}\right)^2, \tag{A.4}$$

Insertion thereof in equation (A.3) gives the following.

$$V_i \simeq \frac{4\pi}{3}r^3\left[1 + \frac{3\Delta r}{r} - 1\right] \tag{A.5}$$

This can then be simplified to

$$V_i \simeq 4\pi r^2 \Delta r. \tag{A.6}$$

Appendix B

Derivation of the standard deviation of the density for a Poisson distribution

For number counts that follow a Poisson distribution the following holds.

Mean $= N_i$

Variance $= N_i$

Standard deviation $= \sqrt{N_i}$,

where N_i is the number of particles within a bin i. With an increasing number of particles per shell, the assumption of error bars being Gaussian and thus symmetrical becomes increasingly more valid (Fageot et al. 2017). Each data point is given by its mean and its standard deviation, which is $N_i + \sqrt{N_i}$. In order to get the desired units of the density, the mass on a single particle $m_{particle}$ and the volume of a bin V_i are included as follows.

$$\frac{N_i \cdot m_{particle}}{V_i} \pm \frac{\sqrt{N_i} \cdot m_{particle}}{V_i} \tag{B.1}$$

Here, the first term corresponds to the density of a specific bin and the second term is its standard deviation i.e. 1 σ error. Rewriting the expression for σ_i using the fact that $\sqrt{N_i} = \frac{N_i}{\sqrt{N_i}}$, gives the following.

$$\sigma_i = \frac{\sqrt{N_i} \cdot m_{particle}}{V_i} = \frac{N_i \cdot m_{particle}}{V_i} \cdot \frac{1}{\sqrt{N_i}} \tag{B.2}$$

The first term in the last expression corresponds to the density in a bin and therefore,

$$\sigma_i = \rho_i \frac{1}{\sqrt{N_i}} \tag{B.3}$$

From this, it follows that each bin is weighted according to the number of counts within it. As expected, bins with smaller numbers of particles are given less weight.