Gravitational Lensing of Gravitational Waves

Michael James Wright

Submitted in fulfilment of the requirements for the
Degree of Master of Science

School of Physics & Astronomy
College of Science and Engineering
University of Glasgow

April 2020

Abstract

Gravitational Waves have opened an entirely new window into the universe, allowing for the probing of non-luminous matter. Dark Matter has thus far been an extremely difficult subject to study given that until now, it has been very difficult to directly investigate. Gravitational waves allow for solutions to this - one of these is by examining the phenomenon of Gravitational Lensing.

The lensing of Gravitational Waves is marked by one of several profiles dependent upon the mass-density profile of the object that has done the lensing. It is this that allows for the probing of Dark Matter should it be the case that when detected, it is possible to determine if a signal is lensed and if so, by what profile. Thus it is urgently necessary to investigate the detectability of a lensed signal and to examine the possibility of determining which type a signal is.

In this work, a piece of software was developed in the python programming language using the work of Antonio Hererra Martin [1] who created individual scripts to generate the amplification factors for each of the lensing types considered, and a tool called 'bilby' [2] designed to perform nested sampling and to present the many gravitational wave specific tools of the LIGO Algorithm Library Suite (LALSuite) [3] to calculate Bayes Factors comparing each of the models with each of the other models, presenting a quantitative analysis of the ability to distinguish signal types.

This software was initially tested by comparison of those lens profiles that were able to be fully integrated, and proof that those more complex could not be integrated further, comparison between bilby generated evidence and the evidence analytically calculated for a simple toy problem, and by comparison of lensed and unlensed signal types using a Fitting Factor developed, first analysed in the Cao et. al (2014) paper [4]. Each of these tests were able to justify that the software was in a functional state to be used for its primary task.

Performing the main analysis using simulated data yielded that in principle the signals received at detectors are able to be correctly identified as the signal type that they are, however, that this ability is dependent upon the parameters of the lens and the signal itself. There are limits based in, at least, the mass of the lensing object and the distance at which the signal originates from the observatory.

Contents

List of Tables

List of Figures

Declaration

I declare that the work in this thesis is entirely my own, except where otherwise indicated and cited. No part of this thesis has been submitted elsewhere for any other degree or qualification.

Acknowledgements

A thesis is not merely the product of the author, it is the combined result of a number of people offering invaluable assistance, feedback, and distraction. First and foremost, I'd like to thank my mother, Morven, my sister, Laura, and my grandparents Catherine and Peter without whom I would not have been able to undertake the year of research and writing required to produce the document before you.

I would next like to thank my supervisor, Professor Martin Hendry, who has been a wonderful guide in my research and an excellent editor in making this thesis as good as it can be for a reader. I would additionally like to thank him for directing me to this particular area of research, it has been a tremendous pleasure to undertake and has given me a path to follow in the future.

I would like to thank Dr Matthew Pitkin for suggesting I use the more user friendly bilby rather than the full LALSuite to perform the analysis, which made creating the code a much simpler task than it would otherwise have been.

A special degree of thanks goes to Chris - the closest friend one could ask for. Thank you for always being either the kick forwards or the desperately needed distraction that is needed to keep sane in undertaking such a big project. Additionally, equal thanks to Fiona and to the members of the MSc Squad. Each of you has been a delight and a pleasure to be around and has made what could have been a lonely year, an excellently fun one!

Finally, I would like to thank you for reading my thesis. A large amount of effort was put in to making this as readable as possible, and I hope that you find that that effort has paid off and what follows educates you in a pleasant way.

Chapter 1

Introduction

For as long as it has existed, questions about the nature of the world around have been asked by humanity. The search for the answers to these questions has driven the development of culture, religion, philosophy, and, perhaps most importantly for direct answers, science. As the lattermost has developed, some questions have been answered but the continual horizon broadening that has come has also posed new questions that earlier seekers could scarcely have dreamed of. Those answers that have been achieved are also not static, they continuously develop and change as thinking and evidence has evolved, leading to the development of whole fields dedicated to determining which answer is the most likely to be correct.

The study of the universe as a whole is simultaneously a very young and very old field of study. The desire to know not just the nature of things on Earth, but the nature of the celestial has been of fascination to a great number of people. In its oldest forms, this would lead to the development of mythology - folk tales that would evolve into pantheons of Gods and the development of religions. More scientifically, the broad field of astronomy has allowed for the expansion of known horizons from an Earth at the centre of many heavenly realms to the geocentric view of Earth as a planet surrounded by the Sun and all other celestial bodies, to what is now known. That the Earth is simply a rocky life-supporting planet in orbit around the Sun, which is itself part of a galaxy called the Milky Way inside the Local Cluster of galaxies, inside of a Local Supercluster. A place that has no specific significance to the Universe as a whole because the Universe should look the same to an observer looking in any direction at any place within it.

With this knowledge of a vast Universe, there is now modern Cosmology - the study of the large scale structure and evolution of the Universe as a whole. It seeks to learn why the Universe appears the way that it does, how that came to be, and what the ultimate fate of the Universe will be. It does this by theorising models and testing them to find out which ones stand up to the observations that are made using the tools that have been developed. One of the major tools for doing this has been the development of statistics -

the mathematical language by which the abstract quantities of knowledge, of likelihood, of evidence can be made quantifiable. There have been two major schema of viewing statistics coming from viewing its base element - probability, a number between zero and one, - in one of two ways.

The first of these is to view the probability of an event occurring as being the percentage frequency that the event occurs given an infinite number of repeated trials. An event occurring 100% of the time corresponding to 1, an event occurring 50% of the time corresponding to 0.5 and so on. This is called Frequentism, and is of great use in many fields such as medical trials [5]. The second method by which one can define probability is as the quantification of belief about an event which may change given new information [6]. In this case, a probability of 1 corresponds to certainty that an event will occur. This is known as Bayesian statistics, and this has particular value in Astronomy, and Cosmology in particular, because it allows for a more in depth examination of things which cannot be infinitely repeated in a lab.

The way in which statistics can help in the advancement of understanding is in the power of Model Selection. As a question is examined scientifically, it may be the case that there are multiple possible causes for a specific phenomenon or multiple theories as to how a phenomenon works. Statistics allows for the analysis of any data gathered from the phenomenon. Specifically, it allows for a calculation of how likely each model is based on the data. Knowing that information, obviously allows for the selection of the most likely model. These calculations are also continuously evolving - new data can be incorporated after the fact meaning as every new data point or set becomes available, the likelihoods of these models can be continuously rechecked to see how their probability changes.

One major example of a phenomenon that exemplifies the evolution of thinking over many centuries, and today both has some answers that provoke yet more questions is the study of Gravitation. This will be shown in detail below to illustrate how scientific thinking has developed over time, but in brief it has gone from a simple question of why things fall to the ground to the flurry of deep questions about the very nature of spacetime that are associated with the simplicity of General Relativity. This theory, our present best to understand gravity, has also allowed for the opening of an entirely new way of looking out into the night sky.

Until recently, there has only really been one way to examine places other than Earth has been simply to look at them. A statement that encompasses the simplest methods of doing this, as in taking it literally, and the most complex methods of doing this by examining in detail the properties of the electromagnetic radiation received at Earth, or sending telescopes or other probes to the few destinations within the relative ease of modern space-faring technology. However, the development of General Relativity has given a completely different way of looking at the universe, that being looking at Gravitational

Waves, which as of September 2015 are now confirmed to exist and have been directly detected by LIGO and VIRGO. [7].

The extension of the ability to examine the universe to beyond just the electromagnetic radiation that is incident upon the Earth should allow for a deeper investigation into some of the most important remaining questions about the universe - specifically it will allow for a probe of dark matter - a kind of matter that does not interact electromagnetically but is believed to constitute a much larger portion of the total matter of the universe than the luminous matter with which people interact with. The nature of dark matter is of great concern to Cosmology as both the amount and distribution of dark matter partly determine the ultimate fate of the Universe and will also help to constrain the plausible models of how the universe has come to be in the state that it is presently in.

One such method by which the examination of Gravitational Wave data will be able to probe the nature of dark matter is in the phenomenon of gravitational lensing. Predicted by the theory of General Relativity, gravitational lensing is the notion that as spacetime is warped by the presence of massive objects, light, or indeed gravitational radiation that passes by these distortions will get distorted also in a manner similar to the lensing of light through materials.

Precisely how the radiation passing by the distorted spacetime is lensed is dependent upon precisely how that spacetime is distorted i.e. how the mass is distributed in the lensing object. Therefore by examining Gravitational Wave signals that are lensed, if it is possible to determine which model of lensing is most likely from that signal, then it is the case therefore that how the object is distributed can be determined. Objects which are made primarily of dark matter can therefore give indications about the properties of dark matter which thus helps constrain which models of dark matter are plausible.

1.1 Development of Gravitational Theory

1.1.1 Historical to Modern

As previously mentioned, it is instructive to look at the evolution of the theory of gravity as a whole as a means of contextualising how the evolution of each of these three fields has contributed to the development of the other two.

The first question in studying gravity has presumably been asked for as long as anyone has ever cared to wonder about the world around them: that being “Why do things fall to the ground”. The first recorded attempts at answering this question started in Ancient Greece.

At this time, there was a belief in five elements: those being Earth, Wind, Fire, Water and Quintessence which pertained to the heavens. Specifically it was the material of rotating spheres upon which astronomical bodies were embedded. This was the Celestial

Spheres model of Cosmology and is shown in Figure 1.1.1. At the centre of this model of the Universe is the Earth, and Aristotle believed that gravity was the result of elements either being heavy and falling towards the centre of the universe i.e. towards the ground or were "light" and rose upwards through the spheres. As a consequence of this, he believed that heavier objects would fall towards this centre faster than lighter objects [8].

Figure 1.1: Pictorial representation of the Celestial Spheres depicted in *Cosmographia* by Peter Apian (Antwerp, 1539) [9]. Sourced from Grant, E. (1987) "Celestial Orbs in the Latin Middle Ages [10]"

This would be the prevailing belief up until the sixteenth century, at which time the geocentric view of the universe had become increasingly complex in an effort to counter the behaviours that did not support this model. It was challenged by the heliocentric model, which placed the Sun at the centre of the system and the Earth as merely a rotating body around it. Evidence would be gathered for this and against the geocentric model by the invention of one of the most crucial pieces of equipment in observational astronomy: the telescope. Through the use of this new device, Galileo uncovered the existence of the four Galilean Moons of Jupiter Io, Europa, Callisto, and Ganymede shown in Figure 1.1.1.

Galileo would also be responsible for the end of the Aristotle view of gravity as being related to an object's weight as he demonstrated that objects of different weight would fall at the same rate.

This would not be the only major advance in the development of gravitational theory at this time, however, as Isaac Newton would publish Philosophiæ Naturalis Principia Mathematica in which he would outline the foundations of what is now known as Classical or Newtonian Mechanics - those being his laws of motion, his universal law of gravitation, and a mathematical derivation of the up until that point empirical Keplerian laws of planetary motion. Now, gravity is a force which attracts all objects with mass together with a strength that is inversely proportional to the square of their separation and it acts immediately in response to changes in the system [11].

Figure 1.2: Composite of the Galilean Moons of Jupiter. From Left: Io, Europa, Ganymede, and Callisto. Image Credit: NASA JPL.

This view of gravity would hold up as the field of cosmology would undergo similar dramatic expansions in understanding. The universe would be infinite in extent and static in time and there would be a singular structure to this universe - a galaxy called the Milky Way. However, in 1755, Immanuel Kant in his *Allgemeine Naturgeschichte und Theorie des Himmels (Universal Natural History and Theory of the Heavens)* would postulate that certain objects in the night sky were in fact other galaxies that he called 'Island Universes', so the view was then that the Milky Way was surrounded by these but was still in the centre of the Universe [12]. The heliocentric picture of the Milky Way would then fall apart as Jan Oort would determine that the galaxy's centre was not the Sun but something else which is now known to be the Supermassive Black Hole, Sagittarius A* [13].

It would be at this point in history that the foundations for what would become Bayesian statistics would be laid, starting with the eponymous Reverend Thomas Bayes whose work would inspire the theorem that bares his name [14]. However, the development of the frequentist view of probability, in response to some of the seeming circular reasoning in parts of the classical view of probability, would prove to be the more widely used at this time due to it seeming to be the more objective view.

Newtonian Mechanics, as mentioned, would hold in large part over the next two cen-

turies after Newton. However, there were certain issues with the theory that emerged over this period. The most major of this was the development of the theory of electromagnetism, specifically the Maxwell Equations and the Michelson-Morley experiment failing to detect the Aether [15]. This would lead Albert Einstein in 1905 to publish *Zur Elektrodynamik bewegter Körpur (On the Electrodynamics of Moving Bodies)* [16]. Here, he would lay out what is now known as his theory of Special Relativity in which the speed of light is fixed for every inertial reference frame and the laws of physics are invariant in these frames.

However, the complete overhaul in understanding of gravity would wait another ten years until he published his theory of General Relativity detailed in Chapter 2, in which objects with mass curves spacetime and the curvature of spacetime causes the motion of objects. Gravity therefore is a property of the geometry of spacetime. This remains to the present day, the most complete description of gravity available and has accurately predicted for instance, the advance of pericentre which explained discrepancies in the orbit of Mercury that the Newtonian theory had been unable to account for [17]. General Relativity has been the underpinning of the development of modern cosmology which will be detailed in Chapter 3 and has led to the theory and detection of gravitational waves and gravitational lensing which are the focus of the remainder of this work.

1.1.2 Outstanding Questions

Where frequently in the historical development of gravitational and cosmological theory, competing models have made predictions or assumptions that were mutually exclusive, for instance heliocentrism vs geocentrism, models examined in modern astrophysics frequently make predictions that are similar or at least not readily obvious given the limitations of equipment available to perform observation with. As a result, statistics plays a vital role in modern science allowing the comparison of these models in a more objective and mathematically rigorous ways. Unlike in the more terrestrial fields, astrophysics does not really allow for the repeated trials that perform much of the justification of the frequentist view and as such Bayesian statistics, which is discussed in detail in Chapter 4, has had a renaissance in development and use in recent times. Of particular value, is the odds ratio which is a direct mathematical statement of whether one model or the other is more likely given the data that has been observed.

Both Cosmology and Gravitation are active areas of research that have several questions that are as yet unanswered. For instance, General Relativity is a classical theory having been largely formulated prior to the development of Quantum Mechanics and latterly Quantum Field Theory. The latter field being an effort to unify the other three fundamental forces of nature: the strong and weak nuclear forces and electromagnetism into a combined framework, however, thus far gravity and General Relativity have yet

to be successfully integrated into this framework. Theories specifically dealing with the quantisation of gravity such as String Theory have been developed in an effort to do this though they are as yet incomplete [18].

This lack of information about the unification of gravity with the other forces also prohibits Cosmology from describing the history of the very early moments of the Universe in which the four forces would be yet to separate from one another. Cosmology has further questions yet to answer, such as the determination which of its predictions will apply to the Universe. Additionally, Cosmology has the questions of both Dark Matter and Dark Energy to answer.

Chapter 2

General Relativity

As has been previously mentioned, General Relativity is the most complete theory of gravitation that has been put forward. The central theme of the theory is well described by a quote of John Wheeler's that *"Space acts on matter, telling it how to move. In turn, matter reacts back on space, telling it how to curve"* [19]. In other words, spacetime is curved and that curvature is responsible for the motion of matter. That curvature is itself the result of the mass and energy of matter warping spacetime.

2.1 Special Relativity

2.1.1 Postulates of Special Relativity

The foundations of General Relativity come from the theory that came before, revolutionising mechanics whilst at the same time leaving the over arching Newtonian theory of Gravitation untouched. Classical Mechanics views time and space as two distinct entities which are absolute. It would also view spacetime as completely flat, describable completely by the Cartesian co-ordinates that are most familiar to the average person and in which distance is described in the Pythagorean way as [20]:

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \tag{2.1}$$

where $\mathrm{d}s^2$ is the invariant interval and the three other parameters are the intervals between two considered points in each of the Cartesian axes. The change that Special Relativity would make would be to unify space and time into a combined spacetime - the Minkowski spacetime. In this combination, the concept of distance also includes the distance in time and becomes invariant to the choice of reference frame for any observer moving with constant velocity (detailed further in §2.1.2) and is calculated as follows [21]:

$$\mathrm{d}s^2 = -\left(c\mathrm{d}t\right)^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \tag{2.2}$$

where c is the speed of light This combination is not the only addition that Special Relativity makes, however, it also comes with two postulates concerning reference frames and the speed of light. These postulates are as follows [22]:

1. The laws of physics must be the same in every inertial reference frame.
2. The speed of light in a vacuum, c, is also the same in every inertial reference. This value is 299,792,458 ms^{-1}

2.1.2 Inertial Reference Frames

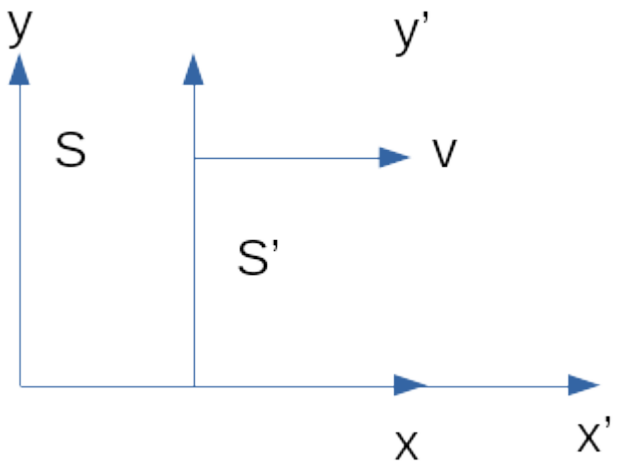

Figure 2.1: 2-Dimensional Illustration of a stationary reference frame $\mathcal{S}$, and a moving reference frame $\mathcal{S}'$ with speed v

The postulates of special relativity both make reference to the concept of an inertial reference frame. An inertial reference frame is one in which there are no unbalanced forces acting upon the entire frame [23]. As a result of that it is a frame that is not accelerating so is either at rest, or is in motion at constant velocity. The postulates state that each of these frames is equivalent, and thus it is possible to make a transformation between them. In Classical Mechanics this transformation is the simple Galilean transformation in which a frame $\mathcal{S}'$ which is moving at velocity v along the x-axis relative to frame $\mathcal{S}$ has only the x-co-ordinate changed by a factor proportional to that velocity. This is given mathematically below [24] and represented in Figure 2.1 above:

$$\begin{aligned} x' &= x - vt \\ y' &= y \\ z' &= z \end{aligned} \tag{2.3}$$

However, in Special Relativity this transformation is replaced with the Lorenztian transformation in which a change in position also causes a change in time and vice versa. This transformation is given by [25]:

$$\begin{aligned} t' &= \gamma\left(t - \frac{vx}{c^2}\right) \\ x' &= \gamma\left(x - vt\right) \\ y' &= y \\ z' &= z \end{aligned} \tag{2.4}$$

where the factor γ is called the Lorentz factor and is given by:

$$\gamma = \frac{1}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \tag{2.5}$$

This is consistent with Newtonian mechanics at lower velocities at which $\frac{v}{c}$ tends to zero and the Lorentz factor as a whole approaches one reducing the Lorentz transformation into the Galilean transformation justifying the apparent accuracy that Classical Mechanics had when considered historically.

2.1.3 Consequences of Special Relativity

Moving from Classical Mechanics to Special Relativity and the associated move from the Galilean to Lorentzian Transformation has consequences beyond the simple change in the value of each co-ordinate. Special Relativity presents the speed of light as a universal speed limit above which no information may travel. This means that the concept of simultaneity only has merit inside of the light cone of an event. Outside of this light cone, events which occur can neither cause nor be caused by the event in question. This means that for instance, should the Sun disappear from existence it would take the approximately eight minutes that light takes to travel to Earth for its orbit to alter due to the distance between them of 1AU [26], whereas when originally proposed by Newton that change should have happened immediately.

The consequence of using the Lorentz transformation over the Galilean Transformation are the phenomena of length contraction and time dilation. The former is the phenomenon that in its a rest frame one may measure the proper length of an object using the means that one might expect. However, should that object be moving relative to the person performing the measurement, the length that will be measured for the object will be shorter by the Lorentz factor i.e [27]:

$$L' = \frac{L}{\gamma} \tag{2.6}$$

The phenomenon of time dilation is a similar phenomenon on the temporal. Consider the scenario of two clocks, one moving relative to the other. An observer in the frame of the clock that is apparently at rest would note that the ticks of the moving clock are slower than those of the apparently unmoving clock. The difference in time intervals are, as is the case for length contraction, a factor of γ [28]:

$$\Delta t' = \gamma \Delta t \tag{2.7}$$

2.2 Prerequisites to General Relativity

2.2.1 Natural Units

Where it is generally accepted that scientific pursuits shall use the the International System of Units [29] , General Relativity prefers to use natural units in order to simplify and make the algebra as elegant as possible. Specifically, it uses a geometrized unit system in which the values of the speed of light, c, and the constant of Gravitation, G, are both set to unity [30]. This removes the need for the in the equations that follow, but it also mathematically allows a representation of the fact that space and time are in many ways indistinct from one another.

The use of $c = 1$ in the natural unitary system and the fact that $c = 3 \times 10^8 \text{ms}^{-1}$ in the SI system means that c operates as a conversion factor between metres and seconds as follows:

$$1\text{s} = 3 \times 10^8 \text{m} \tag{2.8}$$

Similarly, the use of $G = 1$ in the natural unit system and $G = 6.67 \times 10^{-11} \text{m}^3 \text{kg}^{-1} \text{s}^{-2}$ in the SI system combined with the existing conversion between metres and seconds allows G to operate as a similar conversion factor between kilograms and metres as follows:

$$1\text{kg} = 7.426 \times 10^{-28} \text{m} \tag{2.9}$$

2.2.2 Principle of Equivalence

As is noted in Section 2.1, Special Relativity overhauled understanding of mechanics. It did not however, alter gravity, and as a matter of fact does not consider the effects of gravity at all as a requirement of an inertial frame of reference is that a particle in that frame should not move from rest or constant velocity without an applied force. As a result of this requirement, it may seem impossible to apply Special Relativity to a scenario on the Earth, for instance as it may be difficult to separate gravity from the event being examined.

To work around this lack of consideration of gravity, the frame that is considered for an event on the Earth is a free falling frame assuming that the objects within the frame are not too distantly separated. This will meet the requirement for it to be an inertial frame, whilst also being applicable to an Earth based scenario.

The value in making this choice in reference frame with the locality restriction is that in this scenario, gravitational forces become indistinguishable from inertial forces and allow for Special Relativistic predictions to be applied. This is the Principle of Equivalence - that inertial mass (the mass that would be measured by considering this object's resistance to being accelerated) and gravitational mass which could in the most general case be completely separate from one another are in fact not just linked but equal [31]. This partially solves the problem, however, the requirement for the separation between objects to be small in order for the free-falling frame to be applicable, leaving behind one of the motivation for the development of General Relativity.

2.2.3 Vectors, One-Forms, and Tensors

The most familiar mathematical object to the average person are scalars. These are 'just numbers' as a person might say, and are useful in the description of anything that has just a magnitude i.e. it could be used, for instance, to describe how fast something is going but not in which direction it is going. However, to fully look at spaces in a complete mathematical sense in the way that General Relativity does, it is necessary to have a set of objects that can be used to describe the considered space. One way of doing this is to construct a vector space - a group of objects that obey a particular set of rules allowing their use in a description. The rules, or axioms, that are required to form a vector space, V, are as follows, using A and B as two of the objects in the group being used [30] [32]:

- Closure: This is the requirement that addition must exist in the way that is familiar i.e. there must exist an operator '$+$' such that:

$$A + B = B + A \in V \tag{2.10}$$

- Identity: This is a requirement that the addition operator defined above must have an element such that performing the operation with any other element returns that element. In the case of addition this is a zero:

$$0 \in V \text{ such that } A + 0 = A \tag{2.11}$$

- Inverse: This is the requirement that for every element, A , there must exist an element, B, such that performing addition between these two elements will return the identity element as shown below:

$$A + B = 0 \tag{2.12}$$

- Scalar Multiplication: This requirement is that the elements in the set, there must be another element corresponding to the multiplication of any scalar and that element and additionally that multiplying an element by the scalar '1' returns the element:

$$\begin{aligned} \forall A \in V, a \in \mathbb{R} \\ aA \in V \\ 1A = A \end{aligned} \tag{2.13}$$

- Distributivity: This is the requirement that scalar multiplication is distributive. This means that if the result of an addition is multiplied by a scalar as above, this is equivalent to multiplying the two elements by the scalar prior to addition and then performing the addition:

$$a(A + B) = aA + aB \tag{2.14}$$

The easiest example of a complete vector space are the objects simply termed vectors that are perhaps the next most familiar set of mathematical and geometrical objects. These in addition to being able to describe a quantity's magnitude are able to give that quantity a direction. This allows it to be used to describe quantities such as the forces upon an object. Perhaps, slightly less familiar but similar objects that also fulfil the criteria are a set of objects called one-forms. These are also able to describe objects with magnitude and direction and are in many ways almost identical to vectors and when contracted with them, they produce a single scalar. To view both of them in the same space, one may think of a vector as an arrow pointing in a direction with a specific length that represents the magnitude and one may think of a one-form as being a set of parallel planes with the direction being given by that perpendicular to the planes and the magnitude being represented by the spacing of the planes. This is shown below in Figure 2.2.3. In regards the mathematical notation, a vector is represented by $\bar{V}$ and a one-form is represented by $\tilde{p}$.

However, vectors and one-forms by themselves are insufficient to describe all of the quantities and properties of spacetime and matter that need to be considered to fully

Figure 2.2: Pictorial representation of a one-form $\tilde{a}$ and a vector $\bar{b}$. The one-form can be visualised as a series of parallel planes, and the vector as an arrow pointing in a direction

discuss General Relativity. For this purpose, a more general kind of object is required. The objects that are used for this are called tensors. These are a set of objects that like vectors and one-forms form a vector space and thus obey the axioms given above. They are able to be constructed as a combination of vectors and one-forms.

Tensors are given a rank in the form $\binom{M}{N}$ and an $\binom{M}{N}$ tensor may be formally defined a function that is linear in each of its arguments that will take M one-forms and N vectors and map them to an element of the reals. As might seem obvious from this definition, both vectors and one-forms are ranks of tensor, specifically ranks $\binom{1}{0}$ and $\binom{0}{1}$ respectively. Scalars may also be thought of as a rank $\binom{0}{0}$ tensor, requiring neither a vector nor one-form to become a real number.

As is mentioned above, the rank of a tensor indicates how many one-forms and vectors is needed to map to a real number, and a one-form requires a single vector and vice-versa. There is no requirement that contracting a particular vector with a particular one-form would produce the same number as contracting that one-form with that vector, however, this is a restriction that is made in the context of General Relativity:

$$\tilde{p}\left(\bar{V}\right) = \bar{V}\left(\tilde{p}\right) \equiv \langle \tilde{p}, \bar{V} \rangle \quad \forall \tilde{p}, \bar{V} \tag{2.15}$$

An example of a tensor might be the Cauchy Stress Tensor, which for a given point in an object describes the stresses upon that point. It takes two one-forms meaning that it is a rank $\binom{2}{0}$ tensors to produce a real number. The one-forms that it takes are the normal of a real or imaginary surface within the object being described and the displacement of that point. The real number that applying both of these one forms to the tensor returns is the magnitude of the work done per unit area on that point.

2.3 Geometric Description of Spacetime

2.3.1 Spacetime as a Manifold

With the tools given in the previous section, it is now possible to create the mathematical object that describes four-dimensional spacetime as a whole. The object that performs this function is called a manifold. A manifold is very simply a set of points with only one requirement placed upon it - that it is possible to define continuous functions on this manifold [19].

From this requirement, it is possible to define a set of functions $\{x_1, \ldots, x_n\}$ called a chart that maps the points of the manifold to a set of real numbers. The value of n determines the dimensionality of the manifold. This is thus simply a co-ordinate system for the manifold. This also means that the manifold is locally flat, or Euclidean [19].

Paths along the manifold are obviously just a continuous sequence of individual points on the manifold. It is possible to define a function called a curve, $\lambda(t)$, that maps from the manifold to this path. Combining that, with the co-ordinate system defined just above constructs the set of functions $\{x^1(\lambda(t)), \ldots, x^n(\lambda(t))\}$ which relate the path to the co-ordinate system. The sole requirement for the manifold was that the functions defined upon it be continuous. The consequence of this requirement is thus that these functions are differentiable with respect to the parameter t [33].

By and large, the mathematics of working with the manifold are not too dissimilar to the usual systems used in completely flat spaces with which, for instance, Special Relativity are defined. However, that each point on a manifold may have different co-ordinate systems complicates certain matters such as the concept of the subtraction of two points, which due to the definition of differentiation as [33]:

$$\frac{\mathrm{d}f}{\mathrm{d}h} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{2.16}$$

complicates the construction of differentiation in this context. To bypass this issue, the concept of parallel transport comes into play. On infinitesimally small distances, it is possible to define two vectors that are parallel to one another, and by repeating this it is possible to transport a vector from one point to another and hence define the subtraction between two points and allow for covariant differentiation. Covariant differentiation then reduces to partial differentiation inside of Local Inertial Frames hence justifying the continued use of the mathematical methods of Special Relativity in those frames whilst also allowing for expansion into non-inertial frames.

2.3.2 The Metric of Spacetime

For any kind of space that one could construct, the most major geometrical concept that needs to be described is that of the distance between two points in that space. Describing spacetime as a four dimension manifold as described above does not immediately give a method of doing this. To do this, a rank $\binom{0}{2}$ can be defined to take the vectors describing two points and describe the distance between them. This tensor is called the metric, and defines spacetime intervals between a point, A, described by $\{x^0, x^1, x^2, x^3\}$, and a point B described by $\{x^0 + \mathrm{d}x^0, x^1 + \mathrm{d}x^1, x^2 + \mathrm{d}x^2, x^3 + \mathrm{d}x^3\}$ as follows [19]:

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \tag{2.17}$$

where $g_{\mu\nu}$ are the components of the metric tensor. $g(\bar{\cdot}, \bar{\cdot})$. There are few constraints necessarily on the metric but the requirement that the manifold be locally flat means that it must be possible that the metric be reducible to the Minkowski metric, $\eta_{\mu\nu}$, which produces the interval equation given in Equation 2.2, hence [34]:

$$\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1) \tag{2.18}$$

In addition to the Minkowski metric, some of the most commonly used metrics in General Relativity are: the Schwarzschild metric which describes the space around a spherically symmetric massive object such as a star or non-rotating black hole [35], the Kerr metric which describes the spacetime around a rotating black hole [36], and the Friedmann-Lemaître-Robertson-Walker (FLRW) metric which is commonly used in Cosmology [37] - this latter metric is described in more detail in the next chapter.

The metric is also key to determining how a choice of basis vectors behaves inside of a given space described by that metric. It is not necessarily the case that a basis vector is unchanging over the space that it is being applied to. For a vector, $\bar{V}$, covariant differentiation is described by:

$$\frac{\partial \bar{V}}{\partial x^\beta} = \frac{\partial V^\alpha}{\partial x^\beta}\bar{e_\alpha} + V^\alpha \frac{\partial \bar{e_\alpha}}{\partial x^\beta} \tag{2.19}$$

where the second term describes how the basis vectors change over the space. The partial derivative in this term is a vector meaning that it can be broken into a combination of the basis vectors with coefficients:

$$\frac{\partial \bar{e_\alpha}}{\partial x^\beta} = \Gamma^\mu_{\alpha\beta}\bar{e_\mu} \tag{2.20}$$

where the coefficients $\Gamma^\mu_{\alpha\beta}$ are called the Christoffel symbols and they encode how the basis vectors change over the space [30]. Given that they are dependent upon the space and the space is described by the metric, it is thus the case that Christoffel symbols can be

described purely in terms of the metric [32]:

$$\Gamma^{\mu}_{\alpha\beta} = \frac{1}{2} g^{\sigma\mu} \left(g_{\sigma\alpha,\beta} + g_{\sigma\beta,\alpha} - g_{\alpha\beta,\sigma} \right) \tag{2.21}$$

2.3.3 Geodesics

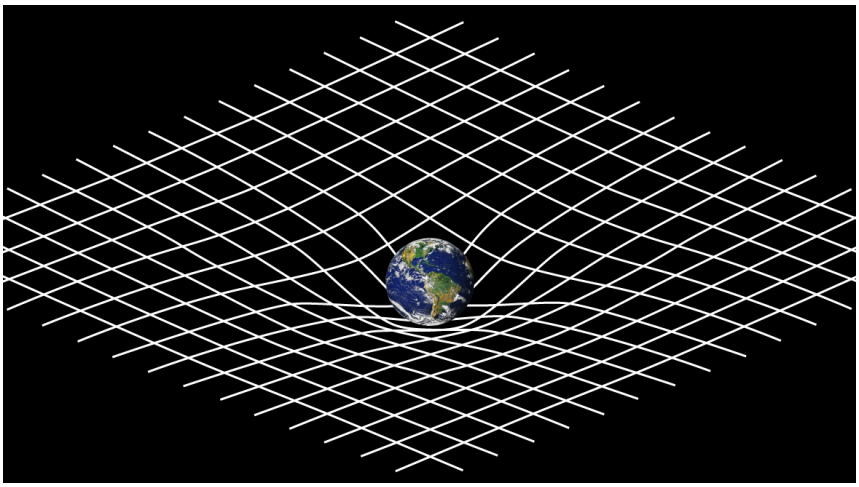

Figure 2.3: An illustration of geodesics and the curvature of space. Away from the curving effects of mass, the geodesic is a straight line but nearer the mass geodesics curve. Image Credit: NASA

The shortest distance between two points is a straight line. This seemingly universally true statement is actually a consequence of the flat Euclidean geometry that describes most human scale concerns. In non-flat space, a truly 'straight line' will no longer be the shortest distance between two points. The more general concept that describes the shortest distance in this case is called a geodesic and is described by [38]:

$$\frac{\mathrm{d}^2 x^{\mu}}{\mathrm{d}\lambda^2} + \Gamma^{\mu}_{\rho\sigma} \frac{\mathrm{d}x^{\rho}}{\mathrm{d}\lambda} \frac{\mathrm{d}x^{\sigma}}{\mathrm{d}\lambda} = 0 \tag{2.22}$$

where λ is the proper time. In the case of flat Euclidean geometry, the Christoffel symbols are all zero which reduces the expression to the first term equalling zero which is just the equation of a straight line.

The separation of two particles, ξ, in gravitational fields tends to change. This change is called geodesic deivation. Describing an interval in spacetime means that the object representing it is a vector. In the case of these two particles being at a certain height, $z(t)$, above the Earth and free-falling towards it, the height and separation are proportional to one another and thus the second derivatives of both of these quantities are proportional to one another. The acceleration of the height is caused by the force of gravity. Inserting the gravitational potential into this proportionality and rearranging yields [33]:

$$\frac{\mathrm{d}^2\xi}{\mathrm{d}t^2} = -\xi\frac{GM}{z^3} \tag{2.23}$$

Thus the presence of a gravitational field or indeed the curvature of spacetime which General Relativity argues are equivalent are indicated by the acceleration of geodesic deviation and also that particles in freefall follow the geodesics of the spacetime which they are in. This applies equally to light as well as massive particles.

2.3.4 Curvature and the Riemann and Ricci Tensors

The curvature of a spacetime is merely a property of that spacetime and like the other properties of spacetime in General Relativity it can be described by a tensor. The tensor that does so is called the Riemann-Christoffel, (or just Riemann), tensor. Given that curvature is the amount that a space deviates from flatness, it is perhaps unsurprising that the components of the Riemann-Christoffel tensor are defined by the Christoffel symbols [39]:

$$R^{\mu}_{\alpha\beta\gamma} = \Gamma^{\sigma}_{\alpha\gamma}\Gamma^{\mu}_{\sigma\beta} - \Gamma^{\sigma}_{\alpha\beta}\Gamma^{\mu}_{\sigma\gamma} + \Gamma^{\mu}_{\alpha\gamma,\beta} - \Gamma^{\mu}_{\alpha\beta,\gamma} \tag{2.24}$$

As can be intuited, in a flat space which is describable by Cartesian co-ordinates, the components of the Riemann tensor will all be zero. Whilst the Riemann tensor is a complete description of the curvature of a spacetime, it is not the only way of describing curvature. A second object that looks at curvature is a contraction of the Riemann tensor on its first and third indices called the Ricci Tensor [40]:

$$R_{\alpha\beta} = R_{\beta\alpha} := R^{\mu}_{\alpha\mu\beta} \tag{2.25}$$

A third object that also can be used to describe the amount of curvature that a space has in a very concise manner is the Ricci curvature scalar. It is defined as the trace of the Ricci tensor [32]:

$$R = g^{\mu\nu}R_{\mu\nu} = g^{\mu\nu}g^{\alpha\beta}R_{\alpha\mu\beta\nu} \tag{2.26}$$

2.3.5 Energy Momentum Tensor

Thus far, the machinery that has been outlaid has been applied to describing the properties of spacetime. However, as General Relativity is about how spacetime and matter affect one another, the same machinery must also be able to describe the properties of matter. In Newtonian physics, the property most important to gravitation is mass, however in relativity this property is expanded to the more general energy-momentum. Energy-momentum can be described using the eponymous energy-momentum tensor, $T^{\alpha\beta}$, the

components of which for a fluid element describes the flux of the α component of its four-momentum across a surface of constant x^β [38].

The case of a perfect fluid, whilst idealised, is a quite useful approximation of many real-world scenarios. For a perfect fluid, in the general case where the fluid has a four-velocity given by $\bar{v} = \{v^\alpha\}$ the components of the energy momentum tensor are given by [30]:

$$T^{\alpha\beta} = (\rho + P)u^\alpha u^\beta + P\eta^{\alpha\beta} \tag{2.27}$$

where ρ is the fluid's energy density, P is the fluid's pressure, and $\eta^{\alpha\beta}$ are the contravariant components of the Minkowski metric tensor. This can be simplified further by viewing the fluid in the frame that moves along with the fluid called the Momentarily Co-Moving Reference Frame (MCRF) where the tensor is simply [30]:

$$T = \text{diag}(\rho, p, p, p) \tag{2.28}$$

The conservation of energy is one of the cornerstones of physics and this too is represented in the energy-momentum tensor by the requirement that its derivative must be zero meaning that the components of the derivative must too be zero i.e. [41]:

$$T^{\alpha\beta}_{,\nu} = 0 \tag{2.29}$$

2.3.6 Field Equations

With a method of describing spacetime using the two curvature tensors as well as scalar curvature and a method of describing matter using the energy-momentum tensor, it is now possible to construct the relationship between them. To simplify the curvature description, the Ricci tensor and the scalar curvature are combined into a single tensor called the Einstein Tensor which is defined as [41]:

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R \tag{2.30}$$

The derivative of this tensor is restricted by an identity regarding the derivative of the Riemann tensor called the Bianchi identities [42]:

$$R_{\alpha\beta\gamma\delta;\lambda} + R_{\alpha\beta\lambda\gamma;\delta} + R_{\alpha\beta\delta\lambda;\gamma} = 0 \tag{2.31}$$

This requires that:

$$G^{\mu\nu}_{;\nu} = 0 \tag{2.32}$$

which combined with the requirement of conservation of energy also requiring that the derivatives of the energy-momentum tensor be zero, it is easy to see that:

$$T^{\mu\nu}_{;\nu} = G^{\mu\nu}_{;\nu} \tag{2.33}$$

Integrating the expression means that the two tensors are proportional to one another which is the Einstein Field Equation, a group of ten equations that state just as was said at the beginning of the chapter that the curvature of spacetime and the presence of matter are intrinsically linked to one another [32] [17]:

$$G^{\mu\nu} = \kappa T^{\mu\nu} \tag{2.34}$$

2.4 Gravitational Waves

The most simple question that can be asked regarding Gravitational Waves is simply what they are. With the language and mathematics of General Relativity, this question can now be answered. A gravitational wave is an oscillatory perturbation of the metric of spacetime caused by motion of mass resulting in a time changing quadrupolar moment [43] [44]. What precisely this means is discussed in the remainder of the chapter.

2.4.1 Weak Gravitational Fields and Nearly Flat Spacetime

As General Relativity specifies, in the absence of matter spacetime should be flat. Therefore, in the case of very weak gravitational fields, the spacetime should be nearly flat. This can be described by taking the description of flat space and very slightly perturbing the metric with a small additional tensor such that the metric of weak gravitational fields is given by:

$$g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta} \tag{2.35}$$

where η is the Minkowski metric used in Special Relativity and $|h_{\alpha\beta}| \ll 1$ is the amount by which the space is perturbed from simply the Minkowski spacetime. The co-ordinate system that describes such a spacetime is called a 'Nearly Lorentz' co-ordinate system [32].

2.4.2 Field Equations in Weak Field

With a metric, it is then possible to describe the tensors needed to form the curvature side of the Einstein Field Equations. These can all be described in terms of the perturbation to the Minkowski metric rather than the whole metric. They are given below [30] [32]:

- Riemann Tensor:

$$R_{\alpha\beta\gamma\delta} = \frac{1}{2}\left(h_{\alpha\delta,\beta\gamma} + h_{\beta\gamma,\alpha\delta} - h_{\alpha\gamma,\beta\delta} - h_{\beta\delta,\alpha\gamma}\right) \tag{2.36}$$

- Ricci Tensor:

$$R_{\mu\nu} = \frac{1}{2}\left(h^{\alpha}_{\mu,\nu\alpha} + h^{\alpha}_{\nu,\mu\alpha} - h^{,\alpha}_{\mu\nu,\alpha} - h_{,\mu\nu}\right) \tag{2.37}$$

  where $h \equiv h^{\alpha}_{\alpha} = \eta^{\alpha\beta}h_{\alpha\beta}$

- Scalar Curvature:

$$R = \eta^{\alpha\beta}R_{\alpha\beta} \tag{2.38}$$

- Einstein Tensor:

$$G_{\mu\nu} = -\frac{1}{2}\left(\bar{h}^{,\alpha}_{\mu\nu,\alpha} + \eta_{\mu\nu}\bar{h}^{,\alpha\beta}_{\alpha\beta} - \bar{h}^{,\alpha}_{\mu\alpha,\nu} - \bar{h}^{,\alpha}_{\nu\alpha,\mu}\right) \tag{2.39}$$

  where $\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h$

Considering the case of free space, the energy-momentum tensor would be equal to zero, which means that the Einstein Tensor would also be zero. Using a gauge transformation it is possible to vastly simplify the Einstein tensor, leaving only [30]:

$$\bar{h}^{,\alpha}_{\mu\nu,\alpha} = 0 \tag{2.40}$$

The metric can be used to raise and lower indices , this means that:

$$\bar{h}^{,\alpha}_{\mu\nu,\alpha} \equiv \eta^{\alpha\alpha}\bar{h}_{\mu\nu,\alpha\alpha} \tag{2.41}$$

Combining this secondary form of Equation 2.40 with the definition of the Minkowski metric and exiting natural units for a moment yields [45]:

$$\left(-\frac{\partial}{\partial t^2} + c^2\nabla^2\right)\bar{h}_{\mu\nu} = 0 \tag{2.42}$$

which is easily recognisable as a wave equation with waves travelling at the speed of light, c.

2.4.3 Generating a Gravitational Wave

The above describes the first half of what a gravitational wave is, the latter half comes from considering how a gravitational wave is generated. Radiation being generated requires that

one of the monopole, dipole, and so on being a time changing quantity. For instance in the case of electromagnetic radiation i.e. light, it is generated by the acceleration of charges. The equivalence between moving a positive charge in one direction and a negative charge in the opposite means that the monopole doesn't exist, however, it also means that a time-varying dipole moment is generated, and this results in electromagnetic radiation.

In gravitation, the monopole represents the total mass which cannot change and the dipole cannot change in time due to the requirement of conservation of momentum. The quadrupolar moment is the next order of moment and it *is* possible for this to be time-varying. The requirement for it to be time-varying is that the motion of mass not be spherically symmetric. An example of such a motion is the inspiral of two masses towards one another such as pairs of Black Holes [45] [7].

Chapter 3

Cosmology

Cosmology is the study of the large scale structure and evolution of the universe as a whole. It is the effort to solve some of the largest and most fundamental questions that humanity has; how did the universe begin, why does it appear the way that it does now, and what will ultimately happen to it. The answer most widely accepted to the first question is the Big Bang Theory which gave the universe a beginning and establishes a universe that changes in size Before the Big Bang Model had been a belief that the universe was a static, unchanging thing that was infinite in its extent [46].

3.1 An Overview of Cosmology

3.1.1 A Brief History of the Universe

The Big Bang model of the universe, as is mentioned in the chapter's introduction, posits a beginning to the universe and that the universe has been and is continuing to expand. This beginning is in and of itself a marked departure from the previously popular infinitely aged static model. The nature of the universe in the very first moments is presently unexplainable, but it is likely to have been extremely hot and dense, as all the energy in the universe was in a very small space [47]. Shortly after its beginning the universe underwent a violent and extremely fast expansion of the universe began. This expansion is called cosmic inflation [48].

With the size of the universe increasing, it also began to cool and this would start to allow the formation of the most elementary particles. This soup of elementary particles would then start the formation of what are now the elements that make up everything with which we are familiar. This process is called Big Bang Nucelosynthesis [49], working based on the ideas that as the universe cooled sufficiently it became possible for protons and neutrons to start forming into atomic nuclei and start creating the lightest elements - hydrogen, helium. lithium, and their isotopes such as deuterium or Helium-3 [50]. Thus

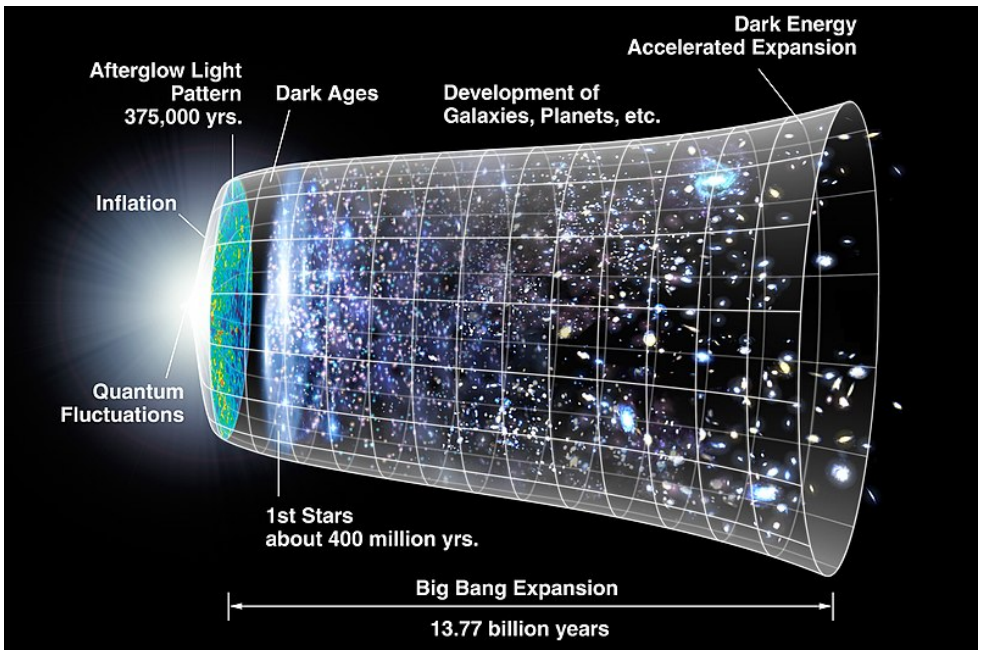

Figure 3.1: An illustration of the evolution of the universe in a hot big bang model - this being the most common model in modern cosmology. Created by the NASA/WMAP Science Team. Original Version: NASA, modified by Ryan Kaldari [public domain] via Wikimedia Commons

elementary matter begins to appear in the universe, however, it was not the dominant thing at the time. In the early development of the universe, it was radiation dominated and that radiation was easily energetic enough that when encountering an atom, any photon was able to ionise it which would release more radiation. As a result, radiation was being constantly scattered on very short timescales meaning that at this time, the universe would be opaque [47].

However, as the universe continued to grow, it would also continue to cool and at approximately 300,000 years after the Big Bang, the photons being emitted would no longer have the energy required to ionise atoms. This time is called the decoupling of matter from radiation or the epoch of last scattering with the latter name coming from the fact that with the formation of neutral atoms and the ceasing of the constant scattering, photons became able to travel much further distances unimpeded, meaning that the universe became transparent. The radiation from this time is still visible as the Cosmic Microwave Background Radiation (CMBR) [48]. During this period, the balance between radiation and matter would also shift from radiation domination to matter domination.

The period of expansion following this is known as the Dark Ages due to the lack of stars or other sources of light. It is a period in the history of the universe that not much is known about [32]. It would last until the universe was approximately 400 million years old, at which point light would shine in the universe with the formation of the first stars. With the universe still being a much higher density than it is today, there were

large amounts of gas available to form each of these first stars and as a result they were much larger, brighter, and shorter-lived than later stars. Their deaths would start the generation of heavier elements, and the creation of galaxies. This cycle of birth and death would continue on, leading to the large scale production of heavier elements and thus the bodies formed from them - asteroids, planets, moons, etc [51]. This process is still continuing in modern times.

At some point in this process, the dominance of the universe has passed once again, this time from matter to dark energy. Little is known about the nature of dark energy beyond that it is believed that it is responsible for the acceleration of the expansion of the universe. This is discussed in more detail in later sections of the chapter.

3.1.2 The Observable and Expanding Universes

The expansion of the universe and the Big Bang hypothesis requires that the universe have a beginning and be of a finite size and age. In addition to that, Special and General Relativity require that the speed of light be finite. Light therefore takes time in order to reach observers, and thus any object that is being observed at significant distance from the observer is not being observed as it is in the present moment but in the past. The finite speed of light also means that there is a finite limit to the amount that can be seen based upon the age of the universe. The amount that can be seen is called the Observable Universe. [52]

As it has been mentioned, it had originally been believed that the universe was infinite in extent and time and that everything was static. What overturned this hypothesis, was the measurement by Edwin Hubble that galaxies had recession velocities i.e they were moving away from Earth [53]. The way that this could be seen was by examining the light that came from these galaxies. Comparing it revealed a difference in wavelength between what was expected had been emitted and what was being received. This difference is called redshift and is defined as follows [52]:

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} \tag{3.1}$$

where λ_{obs} and λ_{em} are the observed and emitted wavelengths respectively. This is related to the velocity that galaxies are receding by, at very small redshift values, by [52]:

$$V = cz \tag{3.2}$$

Comparing the distance and recession velocity of galaxies, Hubble found that these were linked in a linearly proportional manner i.e. that further galaxies were receding more quickly than nearer ones which is an indication that the universe is expanding. This relation has been formalised as Hubble's law [53] [48]:

$$V = HD \tag{3.3}$$

where D is the distance to the receding object and H is called the Hubble constant. Despite the name, the Hubble constant is not actually constant but in fact varies in time. The value of this constant is of great interest to Cosmology as there are several important concepts dependent upon it. It will be discussed in greater detail in subsequent sections.

3.1.3 The Cosmological Principle

Modern cosmological theories have very much overturned the original thought that the Earth was at the centre of the universe; the belief now is that the Earth's location is nowhere particularly special and in fact that nowhere in the universe is a defined "centre". This is formalised in the Cosmological Principle (CP) which states that the universe is homogeneous and isotropic. This means that the universe looks the same to an observer regardless of where in the universe they are (homogeneity) and in which direction they choose to look (isotropy) [32].

It is not immediately apparent that the universe is homogeneous when one looks outwards into the night sky. There would seem to be areas that have a lot in them compared to others, say a bright galaxy compared to an area where there would appear to be nothing. However, if one looks on larger scales, say on the scale of 100Mpc the structures at lower scales smooth out and the universe is very homogeneous and isotropic. This can be seen by looking at for instance the largest scale redshift surveys such as the Sloan Digital Sky Survey (SDSS) [54] [55] .

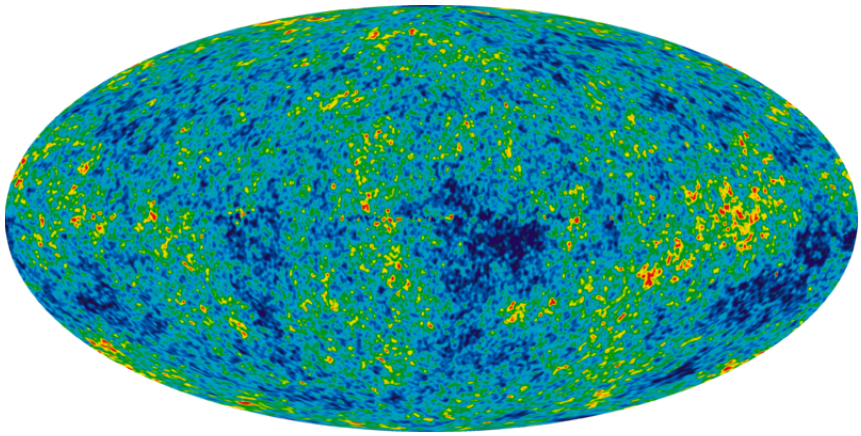

Figure 3.2: All-Sky Map of the Cosmic Microwave Background Radiation (CMBR) created from WMAP data. Image Credit: NASA/WMAP Science Team

The best evidence, however, for the homogeneity of the universe is the Cosmic Microwave Background Radiation. As mentioned in the previous section, the CMBR presents

a picture of the universe at a much earlier time in its evolution. The mean temperature of the CMBR has been measured to be [56]:

$$T = 2.27255 \pm 0.0006\text{K} \tag{3.4}$$

More importantly, however, the variation of the CMBR is on the scale of microkelvin. This lack of variation suggests that at the epoch of last scattering, the universe was as extremely smooth as expected. The small amount of variation that is there, however, then does allow for the development of the structure that is now seen when looking at "smaller" scales [57].

3.1.4 Universal Structure

Figure 3.3: Andromeda Galaxy pictured in ultraviolet by the Galaxy Evolution Explorer. Image Credit: NASA JPL-Caltech

On the smallest universal scales, structures begin with individual stars or stellar remnants such as black holes. These may or may not have solar systems with orbiting planets, asteroids, moons, etc. Stars may also be interacting with one another directly, in for instance binary or trinary systems. The scales of these systems is few tens to a hundred AU.

The next scale of structure is that of individual galaxies - these being clusters of hundreds of millions to trillions of stars, all gravitationally bound to one another. These can come in a variety of shapes such as Spiral, Elliptical, or Irregular. Galaxies frequently have supermassive black holes at their centres, around which the entire galaxy is rotating

in a similar manner to the central star of a solar system. The scales of these are in the tens or hundreds of thousands of light years and are in the $\sim 10^{12} M_{\odot}$ mass range. The galaxy in which Earth resides is called the Milky Way and is a spiral galaxy with a radius of $\sim$ 50000ly.

Galaxies themselves interact in larger scale structures called clusters. These clusters can contain a few tens of galaxies. For instance, the Local Group, in which the Milky Way resides contains at least 50 galaxies, both fully sized galaxies such as the Milky Way and Andromeda, and smaller dwarf galaxies such as the Magellanic Clouds. Larger even, than clusters are superclusters which can contain multiple clusters inside of them - the local supercluster contains the Local Group and the Virgo Cluster. These superclusters are arranged in the largest of universal structures - filaments. Between the threads of the filaments lie much emptier regions of space called Voids - the largest of which are elevated to be termed supervoids. Looking beyond the scale of these filaments and voids is the scale at which the universe's homogeneity becomes apparent.

3.2 Cosmological Models

3.2.1 The Cosmological Metric

General Relativity gives a toolkit by which spacetime can be described in any scenario. Cosmology concerns itself with the properties of the whole of the universe, thus is obviously in need of this toolkit. The most important quantity that needs to be defined in a general relativistic context is the metric of the spacetime being considered. The Cosmological Principle has given two restrictions on the spacetime of the universe - that it be homogeneous and isotropic. A metric which has both of these properties whilst not requiring any other restrictions and thus is widely used in Cosmological considerations is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric which can be written as follows [52]:

$$ds^2 = dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] \tag{3.5}$$

where (t, r, θ, ϕ) are the spherical polar co-ordinate system of coordinates for spacetime. The function $a(t)$ is called the scale factor [52], and it is a dimensionless parametrisation of the expansion of the universe describing how the physical separation of two points changes with time. If it doubles, then the separation has doubled and so forth. The parameter k represents the curvature of the universe. It can take the values of $-1, 0,$ and $+1$ representing negative curvature, flatness, and positive curvature respectively. The precise meaning of these and the implications thereof are described in more detail in a later section.

3.2.2 The Friedmann Equation

How the universe is expanding is very obviously an extremely interesting subject in cosmology. Expansion can be described by the Friedmann equation [58] - one of the central equations in modern cosmological models. In classical Newtonian mechanics it is possible to simply derive it and gain an understanding of its relevance without presenting the full rigour of General Relativity. To do this one can begin from the most important principles in physics - conservation of energy. Specifically, it can be arrived at by considering the gravitational potential energy and kinetic energy of a test particle in a uniformly expanding medium which itself has a certain mass density given by ρ [59].

The kinetic energy of such a particle is calculated, assuming that it is a distance r from the central point of the medium:

$$T = \frac{1}{2}m\dot{r}^2 \tag{3.6}$$

and the gravitational potential is given by:

$$V = -\frac{GMm}{r} = -\frac{4\pi G\rho m r^2}{3} \tag{3.7}$$

using the fact that $M = \rho\frac{4}{3}\pi r^3$ where M is the total mass of the medium. These can be combined to give the total energy of the system which gives an equation that describes the evolution of the separation of two points. This equation is given below:

$$\begin{aligned} U &= T + V \\ &= \frac{1}{2}m\dot{r}^2 - \frac{4\pi}{3}G\rho m r^2 \end{aligned} \tag{3.8}$$

This is not quite the Friedmann equation however. To reach that, the next consideration is that of co-moving coordinates. Co-moving coordinates refer to a coordinate system that changes with the expansion of the universe as opposed to the separation which is based on coordinates that do not alter with time. The translation factor between these is the scale factor, i.e:

$$r = a(t)x \tag{3.9}$$

where x is the co-moving distance. Using this relationship and the equation for the total energy, rearranging, and multiplying both sides of the equation by a factor of $2/ma^2x^2$ then yields the Friedmann equation [32]:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} \tag{3.10}$$

where $kc^2 = -2U/mx^2$. As can be seen with the inclusion of the scale factor, the

Hubble constant, and the constant representing curvature all inside of this equation one can easily see the importance of each in terms of the nature of the Universe at large.

3.2.3 Universal Curvature States

As is mentioned in §3.2.1, the parameter k represents the curvature of the universe, and can take one of three values - these being $-1, 0$ and $+1$. Strictly speaking, it can take any value, but these are rescaled to these particular three values with the scaling absorbed into the scale factor. These values represent the three possible types of curvature [32] [52] [48]:

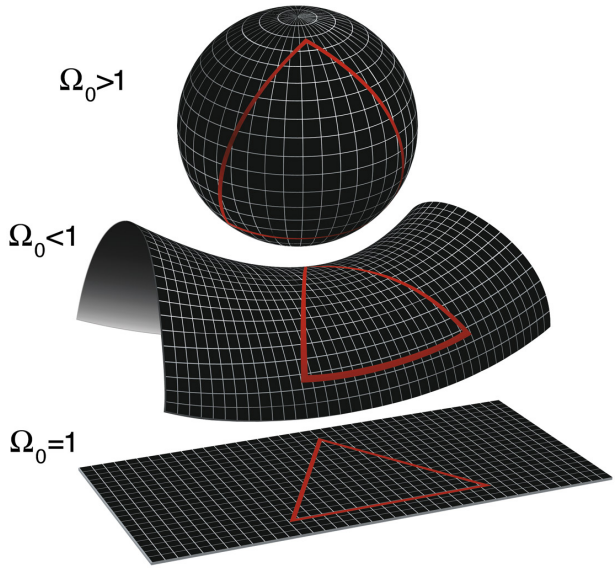

Figure 3.4: Pictorial representation of the three states of curvature. From Top to Bottom: Positive Curvature, Negative Curvature, No Curvature (Flat). Image Credit: NASA/WMAP Science Team

- Negative Curvature ($k = -1$): Negative curvature is best visualised by thinking of a saddle shape. In this geometry, parallel lines will diverge from one another, the angles internal to a triangle add to less than 180°, and the circumference of a circle is greater than the usual $2\pi r$. Universes which have this kind of curvature are called open universes.

- Flat ($k = 0$): Flat geometry obeys the usual Euclidean rules of geometry - parallel lines never meet but stay at precisely the same distance apart from one another, the angles internal to a triangle must add to precisely 180°, and the circumference of a circle is exactly $2\pi r$. Universes with this geometry are simply called flat universes.

- Positive Curvature ($k = 1$): Positive curvature is most well imagined by visualising a spherical type shape. In this regime, parallel lines will all eventually meet, the angles in a triangle will add to more than 180°, and the circumference of circles is given by

less than $2\pi r$. Universes with this kind of geometry are called closed universes - so named because as a consequence of the geometry, if one travels in a 'straight' line long enough, they will return to their point of origin.

3.2.4 Density and Fate of the Universe

One of the major questions of cosmology is what will be the ultimate fate of the universe. The answer to this most profound of questions lies in the density of the universe. The force of gravity is an attraction between every thing that interacts gravitationally - trying to pull together. This force thus opposes the continued expansion of the universe, thus the question of which will win out depends on whether there is enough gravitationally interacting material in the universe to halt the universe's expansion and cause it to recollapse.

The density of the universe required to collapse it is called the critical density of the universe. It can be calculated by looking at the Friedmann equation and considering what density is required to make the universe flat - i.e. to make $k = 0$ [52]. Making this condition and rearranging yields [60]:

$$\rho_c(t) = \frac{3H^2}{8\pi G} \tag{3.11}$$

Using the standard values of each of these constants in the equation, which are $H \approx 100h\text{kms}^{-1}\text{Mpc}^{-1}$, where h is a dimensionless quantity that is used to signify the change from that value [61], and $G = 6.67 \times 10^{-11}\text{m}^3\text{kg}^{-1}\text{s}^{-2}$ [62] gives a critical density of the universe of [60]:

$$\rho_c(t_0) = 1.88h^2 \times 10^{-26}\text{kg m}^{-3} \tag{3.12}$$

For reference, this equates to about 10 hydrogen atoms per cubic metre. The fate of the universe depends exclusively on the relationship between the actual density of the universe and this critical value.

- If $\rho < \rho_c$: this means that the expansion of the universe will win out over gravitation and thus the expansion will continue forever.
- If $\rho = \rho_c$: this means that the expansion of the universe will *just* stop. This means that the expansion rate will reach zero in the limit of infinite time.
- If $\rho > \rho_c$: this means that gravitation will win out over the expansion of the universe and at some point it will reach a maximal size at which the expansion rate will be at zero and then begin to recollapse.

3.2.5 Dark Energy and the Cosmological Constant Λ

The section above discusses the notion that the expansion of the universe is constantly battling against the force of gravity trying to collapse everything back into a single point. It would reasonably be the case that even in the case where the expansion is sufficiently strong that it would defeat gravity, that the effect of gravity should slow the expansion of the universe. The acceleration of the scale parameter of the universe can be given by the so-called acceleration equation [52]:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) \tag{3.13}$$

This allows the direct definition of a parameter, q_0, that describes the acceleration of the universe. This is given by [63]:

$$q_0 = \frac{\ddot{a}(t_0)}{a(t_0)}\frac{1}{H_0^2} \tag{3.14}$$

Using the acceleration equation and the definition of the critical density of the universe, this then means that the deceleration parameter is given by:

$$q_0 = \frac{4\pi G}{3}\rho\frac{3}{8\pi G\rho_c} \tag{3.15}$$

Measurements of the deceleration parameter were one of the most surprising developments in astrophysical history as they suggest that the value of q_0 is negative meaning that despite the effect of gravity, the expansion of the universe is *accelerating* [63]. The phenomenon that causes this acceleration is called *dark energy* . It is so termed because it is not fully understood, but the most common explanation is by a modification of the Friedmann equation [48]:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \tag{3.16}$$

This new parameter Λ is called the Cosmological constant. Principally, the Cosmological constant can be positive or negative, however, if it is positive that allows it to act as an effectively repulsive force and thus overcome gravitation allowing for the measured acceleration [63]. It is additionally possible that the Cosmological Constant acts like the Hubble Constant in the sense that it varies over time, allowing for its contribution to universal density to vary over time. This is determined by the value of a parameter w that relates the pressure and density of Λ - should this parameter be precisely -1 then the effect of the parameter is static through all of time, otherwise the contribution given by the cosmological constant will vary [52].

3.2.6 Dark Matter

When examining the rotation curves of galaxies i.e. comparisons between the speed of rotation and the distance from the centre of the galaxy, the expected fall off in rotational velocity does not occur. The expectation for this fall off came from the distribution of luminous matter in the galaxy, i.e. central bulges and larger flatter discs with the objects in the galaxy orbiting the centre in accordance with the Keplerian Laws.. The observations have yielded a relation that seems to stabilise at larger distances from the centre. An explanation for this behaviour is given by assuming that galaxies are surrounded by large halos of mass coming from a form of matter that cannot be directly seen [64] [65].

Because it cannot be seen, and the nature of this matter is as yet an unsolved problem, it is labelled dark matter. The belief is that it is matter which interacts gravitationally as expected but only interacts weakly with the electromagnetic force, hence not giving off EM radiation. The mass of the particles which make it up determines the speed at which this kind of matter travels. If it is lighter, it is faster, and should it be sufficiently fast as to be relativistic then it is termed 'hot' dark matter (HDM). If it is heavier, and thus not travelling at relativistic speeds then it is considered 'cold' dark matter (CDM). The Standard Model of Cosmology takes the approach that dark matter is more likely to be cold [65].

As is the case where the nature of something is as yet unknown, there are several possible explanations for *what* dark matter is. Two of the most popular theories are that of the Weakly Interacting Massive Particle (WIMP), and primordial black holes. The former case posits that dark matter is the result of a new fundamental particle that would be part of a super-symmetric extention the Standard Model of Particle Physics [66]. Primordial black holes are a possible result from density fluctuations in the early universe allowing for gravitational collapse into black holes [67].

An additional possibility does exist for the nature of dark matter, and that is the Scalar Field Dark Matter (SFDM) model. This resulted from the proposal of the axion as being the source of dark matter. In this model, the fact that an oscillating scalar field can mimic the behaviour of matter is used. Using this behaviour, similar effects to those of the Cold Dark Matter model can be replicated, specifically, the evolution of cosmological density [68]. Further phenomena that this model of dark matter allows are the acoustic peaks of the CMBR [69] , and the scale of sub-structure in the large scale structure of the universe [70].

3.2.7 Balance of the Density Contributions

The three contribution to the universal energy density can be described in terms of the amount of the contributing factor relative to the critical density. This is summarised

with the three Omega parameters, Ω_m, Ω_k, and Ω_Λ for matter, curvature, and dark energy respectively with:

$$\Omega_m + \Omega_k + \Omega_\Lambda = 1 \tag{3.17}$$

The contribution from matter can be further split to show the contributions of luminous (or baryonic) and dark matter respectively (Ω_B and Ω_{dm} respectively). The data from the Planck survey suggests values for these parameters as follows [71]:

Parameter	Constraint
$\Omega_{dm}h^2$	0.1198 ± 0.0026
$\Omega_B h^2$	0.02207 ± 0.00027
h	0.673 ± 0.012
Ω_m	0.315 ± 0.017
Ω_Λ	0.685 ± 0.017

Table 3.1: Cosmological Parameters both given and derived from the 2013 Planck Data [71]

This suggests that the universe is 4.9% luminous matter, 26.8% dark matter, and 68.3% dark energy [52]. As a result of the large amount of dark matter in the universe, particularly compared to luminous matter, understanding the nature of dark matter is a high priority to fully understand the universe.

Chapter 4

Bayesian Inference

As has been previously stated, statistics is the mathematical formalism of analysis and data. It allows for the quantification of more abstract statements about how data compares to theory. The base object of this formalism is probability - the mathematical statement of how likely something is or is not to occur. There are multiple ways of defining precisely what probability is - and each of these leads to a different way of looking at statistics as a whole. One of these is Bayesian statistics which is detailed in the following sections.

4.1 Overview of Statistics

4.1.1 Definition of Probability

Probability is the foundation of all statistical examination and as stated above, is broadly the object describing how likely something is to occur and there are multiple ways of precisely defining this. However, each of the descriptions of probability share how probability is formalised. It is a number between 0 and 1, with the former describing that an event will not occur and the latter describing that an event *will* occur. There have been three major definitions of what probability is - the Classical Definition, the Frequentist Definition, and the Bayesian Definition.

The Classical definition, defines probability as a consequence of the Principle of Indifference - which is the notion that in the absence of sufficient evidence to the contrary when an event may have several results, each of these results is equally likely. An example of this being for instance, dice which unless loaded, should land on each of their sides an equal number of times. Classical probability is thus defined as the ratio of the number of times that that result occurs compared to the number of times that it would occur in the case where each event is equally likely to occur [72]. This definition, however, can be criticised as being circular [73] and as a result, other definitions of probability have come to the forefront.

The Frequentist definition of probability arises out of an evolution of the Classical definition in an effort to remove the circularity criticism by the inclusion of empiricism. In the Frequentist definition, probability is approximated by the relative frequency of the result of an event. That is, that for a series of n trials, the approximation of probability of result x is given by [74]:

$$p(x) \approx \frac{n_x}{n} \tag{4.1}$$

where n_x is the number of times result x occurs. This, however, is only an approximation of probability. The true frequentist definition of probability is the extension of this to the infinite limit of n. In this case, this relative frequency of occurrence will converge to the true probability of result x occurring i.e [5]:

$$p(x) = \lim_{n\to\infty} \frac{n_x}{n} \tag{4.2}$$

This definition of probability has been in wide usage since the nineteenth century and the development of frequentist statistics using this definition has been in wide usage across science since its inception since it is easily applicable to scenarios where repeated trials are easily performed - for instance in medicine and other empirical disciplines. However, the difficulty in the application of frequentist statistics is in scenarios where repeated experiments are difficult or impossible - either due to time-limitation or cost, etc - which is a common occurrence in astrophysical phenomena.

In such scenarios, where frequentist statistics are difficult to apply, there is the third and final definition of probability. This is the Bayesian definition of probability in which probability represents the amount of knowledge or 'plausibility' of the event in question [5]. What differs in Bayesian probability compared to frequentist probability is that in frequentist statistics all statements - for instance, the value of a parameter - are either wholly true or wholly false, statements in the Bayesian view have probability assigned to them, marking the difference between the frequentist view that "for a number of repeated trials, we would expect that x% of the results would be in this range" and the Bayesian view that "we are x% confident that the value of this parameter is in this range".

Bayesian probability was originally conceived by the Reverend Thomas Bayes in 1763 [14], which makes it an older branch of statistics than frequentism, however, it was largely ignored in favour of the development of frequentist statistics due to the belief that the focus on empiricism employed in the latter made it more appropriate for usage in scientific pursuits. In modern times, however, with many fields of research in which Bayesian statistics have proven to be quite useful, it has enjoyed a resurgence in usage and development.

4.1.2 Boolean Algebra

Probability is the basest unit of statistical mathematics. Assigning the extremes of its value to true and false, allows the usage of the framework of Boolean Algebra to develop the mathematics of statistics.

For a proposition, A, the beginning is to assign a symbol to the case of "not A", or the case in which A is false. This symbol is $\bar{A}$. The next step to introduce is the concept of dependence - that is that the truth or falsehood of a statement is dependent upon that of another. For a piece of background information B, the probability of A given the truth of that background information is given by $p(A|B)$. From the combination of both of these, comes the Sum Rule [74]:

$$p(A|B) + p(\bar{A}|B) = 1 \tag{4.3}$$

This is merely a statement of the obvious - that the total probability of A being true or $\bar{A}$ being true given background information B must be 1, since one case must be true.

The next kind of statement to be incorporated into this framework is the concept of joint probability. The probability of A **and** B being true given the truth of background information C is described by $p(A, B|C)$. This allows the definition of the second major rule in this framework - the Product Rule which defines this statement of joint probability to be given by [74]:

$$\begin{aligned} p(A, B|C) &= p(A|C) \times p(B|A, C) \\ &= p(B|C) \times p(A|B, C) \end{aligned} \tag{4.4}$$

An additional concept in Bayesian inference that adds to this framework is the concept of hypothesis space. Bayesian inference is concerned with the truth of one hypothesis or another based upon their plausibility given various background pieces of information. The set of these hypotheses are called a hypothesis space.

There are two kinds of hypothesis space depending upon the nature of the hypotheses being examined. In the case where the competing hypotheses represent discrete states, the space is also discrete. In discrete hypothesis spaces, each hypothesis, H_i, is represented by a probability distribution $p(H_i|D, I)$. The second kind of hypothesis space is the continuous case which applies when the application is towards for instance, the value of a parameter. The value of a parameter, x_0 are considered in the range x to $x + dx$ and the truth of a hypothesis in this scenario can be represented by $p(x_0|D, I)dx_0$ where $p(x_0|D, I)$ is called a probability density function, or PDF. This is defined by [74]:

$$p(x_0|D, I) = \lim_{\delta x \to 0} \frac{p(x \leq x_0 \leq x + \delta x|D, I)}{\delta x} \tag{4.5}$$

For both cases, the belief is that one of the hypotheses is the correct one, meaning

that based upon the information available the correct hypothesis is within the space. The requirement that this place upon the space is one of normalisation. This condition is shown for both kinds of space below [74]:

$$\begin{aligned} \text{Discrete Case: } & \sum_{i=1}^{n} p(H_i|D,I) = 1 \\ \text{Continuous Case: } & \int_{\Delta H} p(H|D,I)dH = 1 \end{aligned} \tag{4.6}$$

4.2 Framework of Bayesian Inference

4.2.1 Bayes' Theorem

The central component to Bayesian inference, is Bayes' Theorem. It is derived by considering the right hand sides of the product rule laid out in Equation 4.4 and is stated as follows [75]:

$$p(A|B,C) = \frac{p(A|C)p(B|A,C)}{p(B|C)} \tag{4.7}$$

Using the hypothesis space framework as described above allows an easier understanding of what each of the probabilities in this equation are. Applying this results in [75]:

$$p(H_i|D,I) = \frac{p(H_i|I)p(D|H_i,I)}{p(D|I)} \tag{4.8}$$

$p(H_i|I)$ is the probability of the hypothesis given the already present background information. It is called the prior and represents the prior knowledge about the hypothesis. $p(D|H_i,I)$ is the probability of getting the experimental data if the hypothesis and the background information both hold. This is called the likelihood. $p(D|I)$ is called the evidence and more fully can be written as $\sum_i p(H_i|I)p(D|H_i,I)$. In part this is a normalisation factor meaning that the left hand correctly normalises as is described in the previous section, but it is also an example of marginalisation which is discussed in greater detail below [74].

The final term in the equation is the left hand side itself, $p(H_i|D,I)$. It represents the state of knowledge about the probability of the hypothesis given the data and the background information i.e. how the experiment has changed the probability. It is thus termed the posterior. Bayes' theorem as a whole thus describes how the statistics of a hypothesis change as every data point or data set is acquired in experiment and observation [74].

The simplicity of this equation gives it great adaptability for performing data analysis for almost any task and gives a well defined starting point for this analysis. This contrasts

with the frequentist approach in which different tasks can require different solutions, leading to a raft of different analytical techniques dependent upon the scenario in question. The secondary difference is also that unlike frequentist techniques, the use of Bayesian inference leads to direct calculation of the probability of the hypothesis.

4.2.2 Marginalisation

Thus far, the discussion has considered hypotheses that consist only of a single parameter but this is rarely the case in real world analysis. In such cases, not every parameter that the hypothesis is dependent upon is actual of value to the analysis being performed. Some are extraneous, and frequently are termed nuisance parameters. Bayesian inference provides a method to deal with these parameters through the concept of marginalisation. To begin the mathematical approach towards marginalisation, it is necessary to start by an extension of the sum rule to a more general state. This is possible in the general case for n parameters, but to illustrate it is simplest to consider only two.

The logical OR is represented in Boolean algebra by the '+' operator and thus the probability of A OR B given C is represented by $p(A+B|C)$. By considering the original sum rule, it is obvious that:

$$p(A+B|C) + p(\overline{A+B}|C) = 1 \tag{4.9}$$

Using the sum and product rules from this starting point and using the fact the logical identity that $\overline{A+B} = \bar{A}, \bar{B}$ it is thus possible to reach the extended sum rule that is needed to proceed to marginalisation:

$$p(A+B|C) = p(A|C) + p(B|C) - p(A,B|C) \tag{4.10}$$

Each of these rules can be extended to larger sets. To create a simple example of marginalisation, consider a model that has two parameters that it is constructed from - a useful parameter θ and a nuisance parameter A. Assuming some background information, I, and that performing an experiment to determine the veracity of the model has yielded some data, D then the joint probability for the model parameters is given by $p(\theta, A|D, I)$. Marginalisation allows for the calculation of the probability of just the parameter θ. [74]

In the simple case where A is a discrete parameter, only able to take one of n values in the set $\{A_1, A_2, \ldots, A_n\}$. If it is assumed that the parameter must be exclusively one of these values then it is the case that [74]:

$$p(A_1 + A_2 + \ldots + A_n) = 1 \tag{4.11}$$

Now using the product rule on the joint probability yields [74]:

$$\begin{aligned} p\left(\theta, [A_1, A_2, \ldots, A_n]|D, I\right) &= p([A_1 + A_2 + \ldots + A_n]|D, I) \times p(\theta|[A_1 + A_2 + \ldots + A_n], D, I) \\ &= 1 \times p(\theta|D, I) \end{aligned} \tag{4.12}$$

In the second term on the right hand side, the probability of θ is given based upon the combination of the set of values the parameter A can take, D, and I being true. The assumption made in equation 4.11 *is* that the parameter A exclusively takes one of these values. As a result, the truth of $[A_1, A_2, \ldots, A_n]$ can be wrapped inside of the background information I, leading to the $p(\theta|D, I)$ term. This is precisely the probability that is being sought - what the data and background says about only the parameter that is considered to be useful.

Considering once again the joint probability, it is possible to rewrite it as, instead of a joint probability of θ and the entire set of possible values for A, a combination of the joint probabilities of θ and each member of the set of values for A, i.e [74]:

$$p(\theta, [A_1, A_2, \ldots, A_n]|D, I) = p(\theta, A_1|D, I) + p(\theta, A_2|D, I) + \ldots + p(\theta, A_n|D, I) \tag{4.13}$$

Combining this with equation 4.12 yields the expression for marginalisation of the case of a discrete variable [74]:

$$p(\theta|D, I) = \sum_i p(\theta, A_i|D, I) \tag{4.14}$$

This can easily be extended to the case of the a continuous nuisance variable by replacing the summation with integration [74]:

$$p(\theta|D, I) = \int p(\theta, A|D, I) dA \tag{4.15}$$

The probability $p(\theta|D, I)$ in the continuous case is called the marginal probability distribution function, and as has been mentioned represents the state of knowledge that results from experimental data and background information on only the useful parameters. This example considers only two parameters, one useful and one nuisance, however, it can easily be extended beyond that to consider as many useful and nuisance parameters are as required by the model. This is an extremely importance advantage that Bayesian inference has over frequentist methods as those methods have no general way of handling nuisance parameters. [74] [75]

The reason that marginalisation is of importance is that increasing numbers of parameters makes computation of the characteristics of the model much more complex and the ability to remove extraneous parameters allows for much speedier calculation, particularly in cases where the integration in the previous equation can be performed analytically.

4.3 Applications and Implementations of Bayesian Inference

4.3.1 Parameter Estimation

It is often the case that models have many different parameters which describe them, and it is almost never the case that the exact values of all of these parameters are known. What is more routine, is that the prediction generated by a model is a range of the constituent parameters. The extent of the range is usually wrapped up by what is known about the model, M, and is formalised in the prior density, $p(\theta_i|M)$ for model parameters θ_i. As is described, most parameters are considered to be inside of a range, if this is assigned to $[\theta, \theta + d\theta]$ then the probability that the true value of θ lies within this range.

For single parameter models with continuous variables, the global likelihood or evidence is given by [75]:

$$p(D|M) = \int p(\theta|M)p(D|\theta, M)d\theta = \mathcal{L}(M) \tag{4.16}$$

In models with multiple continuously defined variables, the global likelihood definition is extended by the addition of an integral for each variable. The definition is thus that the global likelihood for the model is the weighted average of the likelihood of each of the parameters. For the case where the likelihood of the parameter and the prior are both directly assigned, this means that the global likelihood is just a normalisation constant, as described in §4.2.1 and the consequence of that is that the posterior - the probability of the model after data has been acquired is thus proportional to the prior and the likelihood.

This proportionality is the illustration of how Bayesian inference allows for information to be gained about the value of parameters. This process is called parameter estimation - though as opposed to giving individual estimates for each parameter, the posterior is a complete, normalised pdf for the parameters in question. In practice this is done by combining the prior representing the information known before the analysis is performed and the likelihood for the data that is gathered will be produced by the theoretical model.

To summarise the posterior pdf, it is often possible to pick a singular value; the "best-fit" value. This value can then be given appropriate error bars based upon the pdf. The choice of this best-fit value is often either the most probable value of θ, which is given by the posterior mode, or the average value of θ given by the posterior mean [74]:

$$\langle\theta\rangle = \int d\theta \ \theta \ p(\theta|D.M) \tag{4.17}$$

If the posterior mode and posterior mean are vastly different from one another, then this means that the posterior is very asymmetric and cannot easily be summarised by only

a single estimated value. In that case, credible regions become the best way to illustrate the posterior. A credible region is the region in posterior space that the true value of the parameter being estimated has a certain probability to be inside. For probability C, the credible region, R, is given by [74]:

$$\int_R d\theta \; p(\theta|D, M) = C \tag{4.18}$$

4.3.2 Markov Chain Monte Carlo

The process of parameter estimation, calculation of the likelihood function and then multiplication by the prior, is not generally possible through analytical functions. Given that this is not the case, the most complete way to perform this is to create a regular grid in parameter space and sample each point inside of this grid in a systematic manner. However, the disadvantage of this technique is that it is extremely inefficient in higher dimensions. Consider sampling M points for each of N parameters, the total number of points that are sampled in the grid would thus be M^N which quickly becomes large, even for small values for M. To perform Bayesian inference thus, a more efficient method is needed. The choice is that of a stochastic, or random, sampling. This trades the completeness of performing the grid search for computational efficiency.

The most commonly used method of performing this kind of random analysis is Markov Chain Monte Carlo (MCMC), so named for the fact it is a combination of two separate techniques - Markov Chains and Monte Carlo integration. In the latter of these, an integral can be approximated in the following manner.

Supposing the existence of a joint posterior density, $p(\theta|D, I)$. The expectation value of a function of these parameters, $f(\theta)$ can be calculated by performing the integral [74]:

$$\langle f(\theta) \rangle = \int f(\theta) p(\theta|D, I) d\theta = \int g(\theta) d\theta \tag{4.19}$$

Monte Carlo integration proposes that if one uniformly samples n points over a volume V inside of the parameter space of θ then, assuming that the volume is sufficiently large as to cover the area where the function $g(\theta)$ is important to the integral, this integration can be estimated by [74]:

$$\langle f(\theta) \rangle = \int_V g(\theta) d\theta \approx V \times \langle g(\theta) \rangle \pm V \times \sqrt{\frac{\langle g^2(\theta) \rangle - \langle g(\theta) \rangle^2}{n}} \tag{4.20}$$

where

$$\begin{aligned}\langle g(\theta)\rangle &= \frac{1}{n}\sum_{i=1}^{n} g(\theta_i) \\ \langle g^2(\theta)\rangle &= \frac{1}{n}\sum_{i=1}^{n} g^2(\theta_i)\end{aligned} \tag{4.21}$$

The ability to increase n to as high or as low as is needed allows the trading of accuracy for speed. However, the problem with Monte Carlo integration by itself is that the uniform sampling means that the probability of sampling the region with high probability is very low, meaning that the majority of the sampling is spent in wasted regions. Instead of uniformly sampling, the sampling can be done by any process which will return samples from the target distribution, i.e. $p(\theta|D,I)$ in the right proportions. The process which is selected to improve the sampling is, perhaps unsurprisingly, where the Markov Chain comes into play.

In a Markov Chain, each sample θ_{i+1} is dependent upon the previous sample θ_i by an object called the transition probability or kernel, $p(\theta_{i+1}|\theta_i)$. The use of this transition probability means that the samples of θ that it generates have precisely the same probability density as that of the sought after posterior. The way that it works is a two step process: first a proposed value for θ_{i+1}, Θ is chosen from a probability distribution, $q(\Theta|\theta_i)$ which can easily be calculated. The second step is then to decide whether or not to accept the point. This is done on the basis of a ratio, r, which is given by [74]:

$$r = \frac{p(\Theta|D,I)q(\theta_i|\Theta)}{p(\theta_i|D,I)q(\Theta|\theta_i)} \tag{4.22}$$

If $r \geq 1$ then the point is accepted and Θ is accepted as θ_{i+1}. If it is less than one, then the point is accepted with probability r, The means by which this is accomplished is to sample a random variable, U from Uniform(0,1) and comparing to the value of r. If $U \leq r$ then the point is accepted. If the opposite is true, then Θ is rejected and θ_{i+1} is set to be θ_i.

The most widely used implementation of MCMC is done by means of the Metropolis-Hastings algorithm which for an n point chain goes as [76]:

1. Initialise point θ_0
2. For $i = 1$ to n, sample Θ candidate from $q(\Theta|\theta_{i-1})$:
   (a) For candidate Θ calculate r. If $r \geq 1$ accept candidate.
   (b) If $r < 1$, then sample U from Uniform(0,1). If $U \leq r$ then accept candidate
   (c) If $U > r$, then $\theta_i = \theta_{i-1}$
   (d) Increment i

The Metropolis-Hastings algorithm works because the samples of θ are generated with a probability density that converges upon that of the posterior because it is a stationary distribution [76]. The requirements on the Markov Chain for it to converge to a stationary distribution are that it must have the following properties [74]:

- Irreducible: This means that the from all possible starting points, the Markov Chain must be able to reach every state in the target distribution with a positive probability in a finite number of steps
- Aperiodic: This means that the chain is not allowed to oscillate between states in a regular periodic fashion
- Positive Recurrent: This is the requirement that a stationary distribution must exist such that if the initialisation step θ_0 is pulled from this distribution then all the subsequent iterations will also be distributed from that distribution.

4.3.3 Model Selection

Parameter estimation concerns the likelihood of a model given a set of data. It is often the case that there are multiple models available to explain a given phenomenon and thus establishing which model is the most likely. The process of determining which model is the most likely through the auspices of Bayesian Inference is called model selection. In this realm, when the consideration is that one of several models may be the case, then the background information becomes the full combination of all of these models i.e. $I = M_1 + M_2 + \ldots + M_n$ where the operator '+' is the Boolean OR.

Using Bayes' Theorem, it is possible to construct the probability of models in an analogous fashion as to parameter estimation with the posterior in this case being the probability for any given model as a whole, $p(M_i|D, I)$ [74]:

$$p(M_i|D, I) = \frac{p(M_i|I)p(D|M_i, I)}{p(D|I)} \tag{4.23}$$

Once again the denominator merely serves as a normalisation, meaning that the analogy with parameter estimation holds and the posterior for the whole model is proportional to the prior for the model multiplied by the global likelihood of the model which was defined in equation 4.16. In model selection, however, calculation of the direct value of the posterior probability is less informative than the ratio of the posteriors for the models being selected between. This ratio is given by [74]:

$$O_{ij} = \frac{p(M_i|D, I)}{p(M_j|D, I)} \tag{4.24}$$

and is called the odds ratio for models i and j. Inserting Bayes' Theorem into this equation yields [74]:

$$O_{ij} = \frac{p(M_i|I)p(D|M_i,I)}{p(M_j|I)p(D|M_j,I)} \equiv \frac{p(M_i|I)}{p(M_j|I)} B_{ij} \tag{4.25}$$

The first term on the right hand side is called the prior odds ratio. It is frequently the case that the same priors are chosen for both of the models resulting in a prior odds ratio of unity. The second term on the right hand side is called the Bayes' factor and is the ratio of global likelihoods for each models. As shown in §4.3.1, the global likelihood is the result of marginalising over the parameters of the model. The consequence that this has for model selection is a mathematical implementation of Occam's Razor which states that the simplest explanation is most often the correct one.

To illustrate how it is encoded into the Bayes' factor, consider a scenario with two models: M_1 which has a single free parameter θ, and M_0 which fixes the value of θ at a constant θ_0 and thus has no free parameters. The objective is to calculate the likelihoods of each of these models i.e. $\mathcal{L}(M_1)$ and $\mathcal{L}(M_2)$. Beginning with M_1, and assuming that for both of the models the prior choice is a flat with width $\Delta\theta$ then the normalisation for the prior is [74]:

$$\int_{\Delta\theta} d\theta p(\theta|M_1,I) = p(\theta|M_1,I)\Delta\theta = 1 \tag{4.26}$$

This means that

$$p(\theta|M_1,I) = \frac{1}{\Delta\theta} \tag{4.27}$$

The likelihood is calculated by considering the characteristic width $\delta\theta$. This width is defined by [74]:

$$\int_{\Delta\theta} p(D|\theta,M_1,I) = p(D|\hat{\theta},M_1,I) \times \delta\theta = \mathcal{L}(\hat{\theta}) \times \delta\theta \tag{4.28}$$

where $\hat{\theta}$ is the value of θ that has the maximum likelihood. This allows the global likelihood to be approximated as:

$$\mathcal{L}(M_1) \approx \mathcal{L}(\hat{\theta}) \frac{\delta\theta}{\Delta\theta} \tag{4.29}$$

For the second model, the lack of free parameters means that the global likelihood for that model is simple to calculate. It is precisely the same as the probability for the likelihood of M_1 at the value of θ_0, $\mathcal{L}(\theta_0)$. Combining this yields the Bayes factor for these two models [74]:

$$B_{12} \approx \frac{\mathcal{L}(\hat{\theta})}{\mathcal{L}(\theta_0)} \frac{\delta\theta}{\Delta\theta} \tag{4.30}$$

This is a combination of two ratios - likelihood and width. It is impossible for the ratio of the likelihoods on its own to favour the second model since the first model contains the second as a specific case. However, in the ratio of widths, the posterior width $\delta\theta$ is much narrower than the prior width $\Delta\theta$. The effect of this is that the ratio of widths punishes models for having too many parameters that are not necessarily supported by the data. This means that the likelihood ratio must strongly favour the more complex model in order for it to be favoured by the Bayes Factor. To generalise this specific example, global likelihoods can be written as a the combination of the maximum likelihood multiplied by some factor Ω_θ [74]:

$$p(D|M, I) \equiv \mathcal{L}_{\max}\Omega_\theta \tag{4.31}$$

This factor is called the Occam factor - so named because it is the mathematical encoding of the principle of Occam's razor into Bayesian inference. [74]

4.3.4 Nested Sampling

Nested sampling is a computational algorithm that is intended to aid in the task of performing model selection in much the same way that Markov-Chain Monte Carlo can be used to perform parameter estimation. Nested Sampling takes a primary focus on the evidence in Bayes' Theorem as opposed to other cases where the evidence is mostly viewed simply as a normalisation constant. This yields the posterior as well, but as a consequence of the calculation of the evidence rather than as the primary focus. The evidence, Z, is calculated as [77]:

$$Z = \int \mathcal{L} dX \tag{4.32}$$

where $\mathcal{L}$ is the likelihood function and $dX = \pi(\theta)d\theta$ is the element of prior mass. Inserting this into Bayes' Theorem, considering the posterior mass, and rearranging yields [77]:

$$\mathcal{L}(\theta) \times \pi(\theta)d\theta = Z \times p(\theta)d\theta \tag{4.33}$$

As is the case with parameter estimation, performing the necessary integrations to calculate the evidence becomes difficult for models with many parameters due to the high dimensionality that imparts to θ. However, the idea behind nested sampling is that the prior mass X can be acquired from elements in any order. If X is defined such that [77]:

$$X(\lambda) = \int_{L(\theta)>\lambda} \pi(\theta) d\theta \tag{4.34}$$

then X is the cumulative prior mass enclosing all likelihood values that are greater than λ. This means it is a function that decreases from 1 to 0 as λ increases. This means that the evidence becomes a one dimensional integral over the range 0 to 1 [77]:

$$Z = \int_0^1 \mathcal{L}(X) dX \tag{4.35}$$

This can be computed by converting this integration into a weighted sum - accomplished by splitting the likelihood function into a series of weighted columns with weight w_i yielding [77]:

$$Z \leftarrow \sum_{i=1}^{m} w_i L_i \tag{4.36}$$

where $L_i = \mathcal{L}(X_i)$. The idea is to use a series of randomly selected samples of X sorted such that $X_0 = 1$ and that $X_i < X_{i-1}$ and then calculating the values of θ_i. However, it is possible to bypass the need to sort the samples by directly taking values of θ_i with the constraint that $L(\theta_i) < L(\theta_{i-1})$. The evaluation of these sampling points can be done by Monte Carlo methods [74].

Calculation of the evidences is important for the process of model selection, but it is also possible to use nested sampling as a parameter estimation method, because the posterior can be extracted from this process by normalising the sum by the evidence:

$$p_i = \frac{w_i L_i}{Z} \tag{4.37}$$

Chapter 5

Gravitational Waves and Lensing

Gravitational Waves represent an entirely independent method of looking at the universe than what has come before, looking out first in the optical ranges and then in the wider electromagnetic spectrum. Their detection has been the culmination of decades of effort into achieving the very high standards of precision necessary. With the primary challenge of being able to detect gravitational waves at all, one of the next steps will be looking for evidence of gravitational lensing of gravitational wave signals. If detected, lensing may be able to allow gravitational wave observations to probe the nature of dark matter. How gravitational waves are detected, and the theory of how gravitational lensing of gravitational waves may allow this probing are detailed in the following sections.

5.1 Gravitational Wave Detection

5.1.1 Difficulty of Detection

Gravitational Waves were first predicted by Einstein in 1916 [44], and the first detection took place on the 14th of September 2015 [7], almost a full century later. They are produced by any motions of matter that result in a time-changing quadrupolar moment, i.e. motion that is not spherically symmetric [45]. The obvious example of this are pairs of inspiralling objects - such as black holes or neutron stars. Such a scenario is not an uncommon phenomenon, so it would reasonable to ask the question of why it took a full century to move from prediction to observation.

To begin with, consider that gravitational waves are much like other waves and possess the measurable properties of amplitude, frequency, and polarisation. The frequency range of gravitational waves is quite broad, extending from the nanohertz range to the kilohertz range. There are two polarisation states for gravitational waves, these are termed the '+' and '×' or plus and cross states [45]. Which is the case is the result of the orientation of the masses. However, it is the third and final property that is responsible for the difficulty

in detecting them.

Gravitational waves have very small amplitudes. The reason for this is very much a consequence of General Relativity. In Newtonian physics, space is unchanging - this means that space is infinitely stiff which would prohibit the existence of Gravitational Waves. However, General Relativity as discussed prior introduces local curving of spacetime which means that space is no longer infinitely stiff, it is finitely so. This is described by the Einstein equation [45]:

$$\frac{8\pi G}{c^4}\mathbf{T} = \mathbf{G} \tag{5.1}$$

The constant of proportionality between the two tensors in this equation is called the coupling coefficient and using the standard values of these constants yields a value of the order 10^{-43} meaning that whilst space is flexible, it is only barely so. The consequence of this is to make the amplitudes of gravitational waves extremely small. As an example of precisely how small, the strain amplitude, h, for an inspiral can be approximated by [45]:

$$h \sim 1.4 \times 10^{-22} \left(\frac{f}{100\text{Hz}}\right)^{2/3} \left(\frac{\mathcal{M}}{1.22 M_\odot}\right)^{5/3} \left(\frac{10\text{Mpc}}{r}\right) \tag{5.2}$$

where f is the gravitational wave frequency, r is the distance between the source of the gravitational waves and the observer, and $\mathcal{M}$ is a combination of the masses that are inspiralling called the Chirp Mass. This combination is given by [78]:

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \tag{5.3}$$

As can be seen from these ratios and the prefactor, experiments that could be reproduced on Earth using spinning masses would be undetectable. Only the more extreme phenomena that are more commonplace in astrophysics stand a chance of being able to be detected, and even then the level of precision required is extreme. This has been the major stumbling block that had prevented the detection of gravitational waves until recently.

Whilst the strain amplitude formula indicates why detection of gravitational waves is difficult, it also demonstrates one of the major benefits that being able to observe them can bring. With the chirp mass and the frequency, detection of gravitational wave signals allows a direct calculation of distance. This represents a major improvement compared to electromagnetic observations where only a select few types of object are able to have their distances calculated directly and the rest must have their distances calculated based upon those few ‘known’ distances.

5.1.2 History of Gravitational Wave Detectors

The means by which gravitational waves are detected is at once both extremely simple and extremely complex. Fundamentally, all that is necessary to detect gravitational waves are a pair of test masses that are free floating in space and the ability to monitor their separation. The first complication is obvious considering the above section and that is that the change in position of the masses would be extremely slight. The second complication is that almost any disturbance that could be imagined would interfere with the experiment and overpower the gravitational waves.

That the challenges have only recently been overcome, does not mean that attempts had not been made in the past to listen for gravitational waves. As early as the 1960s, the first gravitational wave receivers were constructed by Joseph Weber. These first receivers were simply large masses that could have quadrupolar vibrations excited in them by gravitational waves. He would then go on to design the Weber Bar [79]. These are large aluminium bars that gravitational waves could excite the fundamental longitudinal vibrational mode of. This was later refined into cryogenic resonant mass detectors, which were similar in concept but cooled to reduce thermal vibrations that could prevent the detection of the gravitational waves [79]. There were also even more grandiose notions of using the Earth as a huge scale gravitational wave detector - though it would prove impossible for the Earth to achieve the needed sensitivity [80].

The design of the modern detectors that have achieved successful observations of several objects traces its evolution from the instrument used in another experiment, one of great importance, and that is the Michelson-Morley experiment. This was an experiment to investigate the nature of the electromagnetic spectrum. It was believed at the time that light waves must have a medium through which they propagate and that this medium, called the luminferous æther existed everywhere. Since it existed everywhere, this means that the Earth would be passing through it during its orbit and this would lead to a slight difference in the time it would take for light to travel in different directions. [81]

The apparatus used to perform this experiment is known now as the Michelson-Morley interferometer. It is fundamentally a light source, a beam-splitter, two mirrors and a detector. The light travels into the beam-splitter and proceeds in two seperate beams that are perpendicular to one another, these beams are then reflected by a mirror, return to the beam-splitter, and now recombined pass towards the detector together. This allows for any difference in the path length in the light to be picked up by examining the interference pattern between the two beams of light at the detector.

Modern gravitational wave detectors still use this principle as the foundation of their design though they are, of course, much more sophisticated in engineering than those used in the day of the Michelson-Morley experiment. These are laser interferometers in which the gravitational waves are responsible for minute changes in the distance that of the arms

of the interferometer.

5.1.3 Laser Interferometer Gravitational Wave Detectors

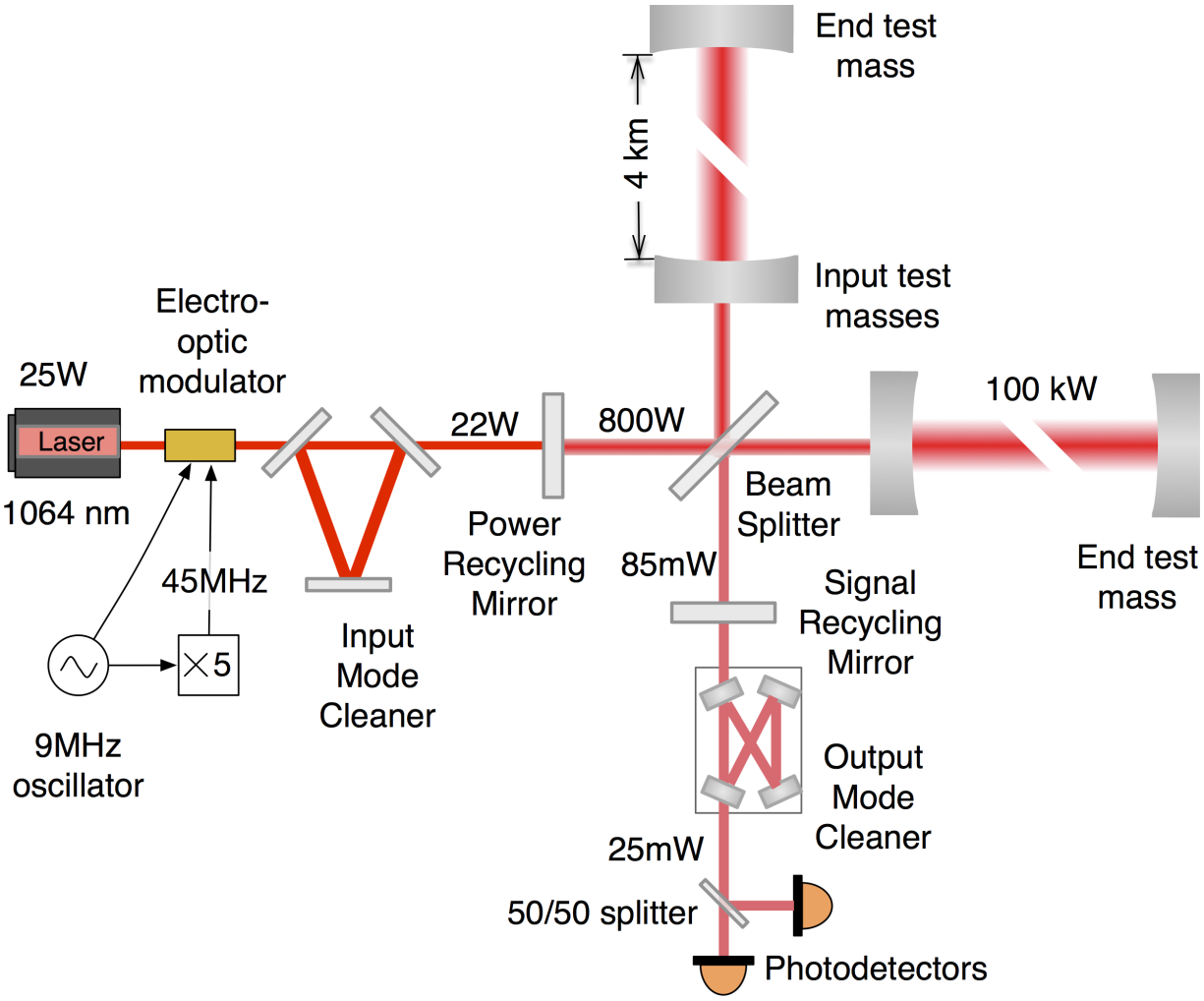

Figure 5.1: Diagram of the LIGO detector set up, a heavily modified variant of a simple Michelson interferometer. Image from [82]

The modern laser interferometer gravitational wave detector is based on the foundation of the Michelson-Morley interferometer. In the Advanced LIGO (Laser Interferometer Gravitational Wave Observatory) detectors, after passing through the beam splitter, each of the beams enters into a Fabray-Pérot cavity between a pair of test masses. This reflects the laser beam back and forwards through the space of the arms multiple times, allowing for a much longer path difference without actually extending the arms. After passing through the cavity and being reflected out, the beams return to the beam-splitter. There are also two additional mirrors in the system, one between the laser and the beam-splitter and one between the beam-splitter and the detection equipment. These are called a power-recycling mirror and a signal-recycling mirror respectively and their function is to

increase the power that is incident upon the beam-splitter and to increase the detector's sensitivity. [83]

These improvements are necessary to achieve the desired sensitivity, as whilst it is theoretically possible to simply build a larger and more actively powerful detector, it is very much impractical. In the Advanced LIGO detector, the power recycling process improves the laser power up to a value of 700 kilowatts, compared to the original power that enters the detector which at maximum is $\sim$ 20 watts [7]. The arm lengths for the LIGO detectors are an already fairly large 4km [83]- the use of the Fabray-Pérot cavities extends this distance to an effective 1120km .

However, this alone only allows for the sensitivity necessary to perform the detection. Were it left only with these improvements, the device would not be able to detect gravitational waves due to the many potential source of error, some of which are easily orders of magnitude higher than the signal. Some examples of these noise sources are detailed below and the figure shows a full comparison of the amplitude spectral densities of all of the noise sources considered for the LIGO detectors.

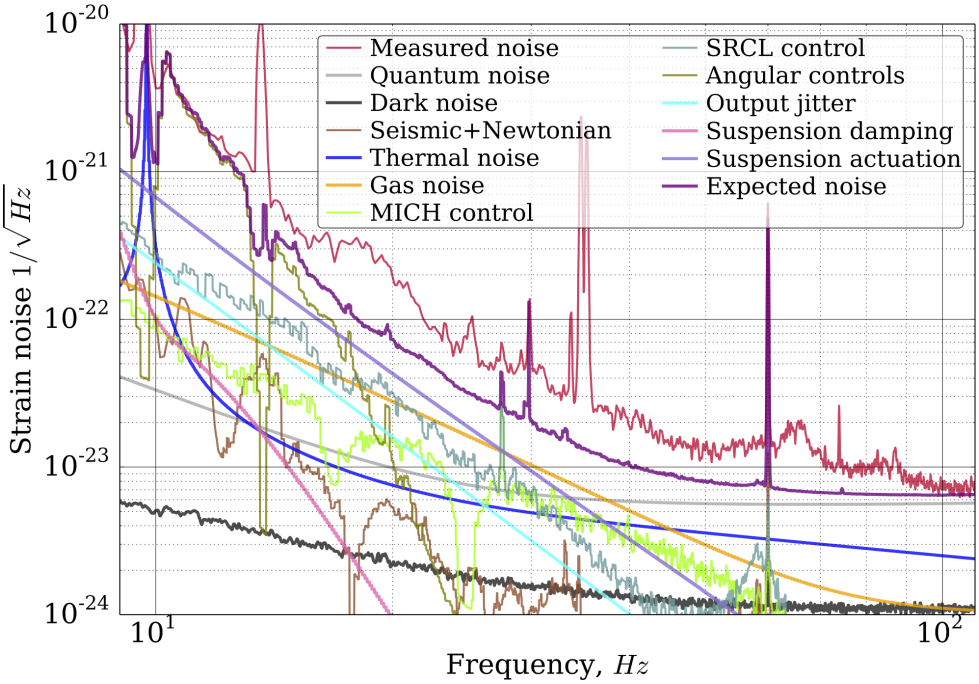

Figure 5.2: Comparison showing the amplitude spectral densities at each frequency for each of the noise sources expected in the LIGO detectors. These are combined in order to form the sensitivity curve of the detector. Image Credit: LIGO Scientific Collaboration, from [82]

The first of these noises sources is called seismic noise and as it sounds, it is the noise caused by the motions in the ground. In terrestrial detectors, it is a large source of noise for the lower end of the frequency spectrum ($<$ 10 Hz) [45], however, is very obviously not an issue in space based detectors. To counter it in ground-based detectors,

two systems of seismic isolation are used, passive and active. The passive system is the suspension of the test masses in a quadruple pendulum system. In each system there are two chains of masses called the Main and Reaction chains. The Reaction chain helps to counteract motions in the Main Chain where the test mass is using a combination of magnetic and electrostatic forces. This quadruple-pendulum system is itself mounted to the active system which monitors various frequencies of geological vibration and uses a series of actuators to counteract any detected motions. [45]

The second noise source is thermal noise. It is one of the fundamental noise limits in gravitational wave detection - this means that whilst it can be reduced, there is no way to completely remove it. Thermal noise is the result of the inherent vibrations in all things due to their temperature - which is why it cannot be removed. In a laser interferometer, it can be divided into three main sources: suspension thermal noise, internal thermal noise to the test masses, and mirror-coating thermal noise [82] [84]. The counter to each of these is to find materials that allow the construction of the device but also minimise the generation of thermal noise. For present gravitational wave detectors, the best material is fused silica, though in the future, sapphire or pure silicon may be used [84].

The third noise source is quantum noise. It is likely to be the limiting factor in the sensitivity of advanced gravitational wave detectors. Broadly, it is the noise that is induced due to the quantum nature of the light source and the photodetector. There are two main factors in quantum noise: radiation pressure noise and shot noise [82]. Radiation pressure noise is the result of quantum fluctuations which exert radiation pressure on the mirrors of the interferometer. This leads to uncertainties in the position of the mirror [82]. Shot noise is the consequence of the nature of light as a photon. Photons are emitted at random moments meaning that at a photodetector, there is a minute variation in the received signal at any given time [45]. This uncertainty is shot noise. As can be seen from their definitions, as is the case with thermal noise, they are impossible to completely remove from the system.

With an understanding of how the noise comes about, it is possible to model how strong it would be at each frequency. This combined modelling yields the sensitivity curve for the detector. This gives the limits of the signals that the detector will be able to detect. The figure below shows a comparison of the sensitivity curves for the LIGO Livingston and Hanford detectors during their first observing runs and their design sensitivity curve.

Combining this signal with the predictions of General Relativity allows the prediction of what signals should look like allowing for the comparison between theory and observation. Figure 5.4 beneath shows the comparison between the signals received at LIGO Livingston and Hanford and these combined predictions.

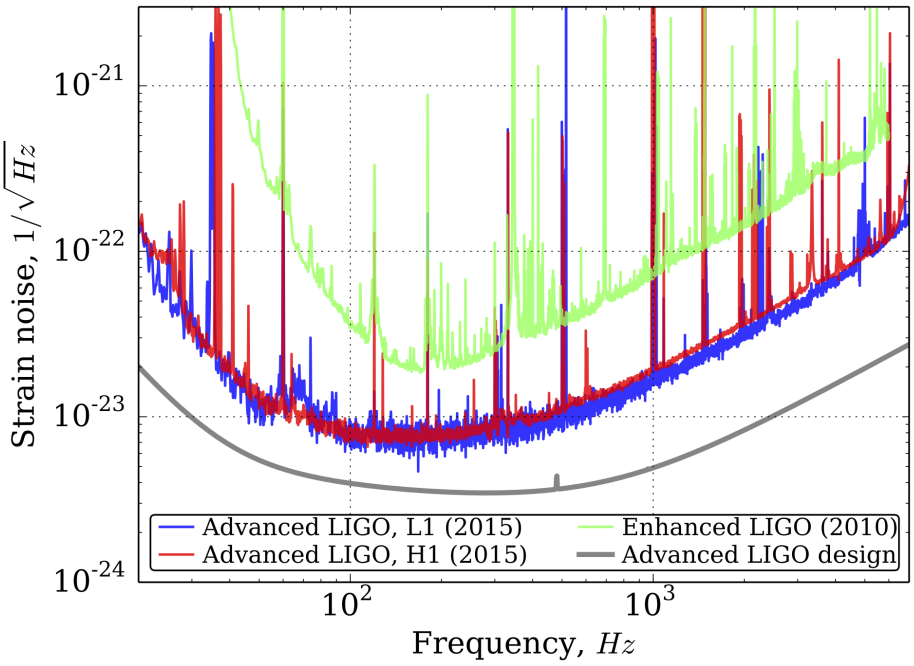

Figure 5.3: A comparison between the sensitivity curves of the LIGO Livingston and Hanford detectors as they were during their first observing run (O1) and their design sensitivity curve. Image Credit: LIGO Scientific Collaboration, from [82]

5.2 Gravitational Lensing

5.2.1 Overview of Gravitational Lensing

Gravitational Lensing is one of the phenomena predicted by General Relativity. Massless objects such as the photons of light travel upon the geodesic paths that were discussed in §2.3.3. General Relativity describes that these geodesics change in the presence of massive objects which curve the spacetime around them. This results in the bending of light, and in certain scenarios this bending can resemble precisely the scenarios of light being bent when passing through lenses. One of the ways that this was verified to be the case, and one of the major tests of General Relativity as a theory was measuring the angle of deflection of light passing close to the Sun in the Eddington experiment of 1919 [85].

The Sun can be reasonably considered a spherically symmetric distribution of mass. This means that the spacetime surrounding the Sun can be described using the Schwarzschild metric, the invariant interval for which is given by [86] :

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \tag{5.4}$$

where M is the mass of the object - in this case the Sun. Following the geodesics of this metric, there will be a path just such that the light's path is curved by the Sun but

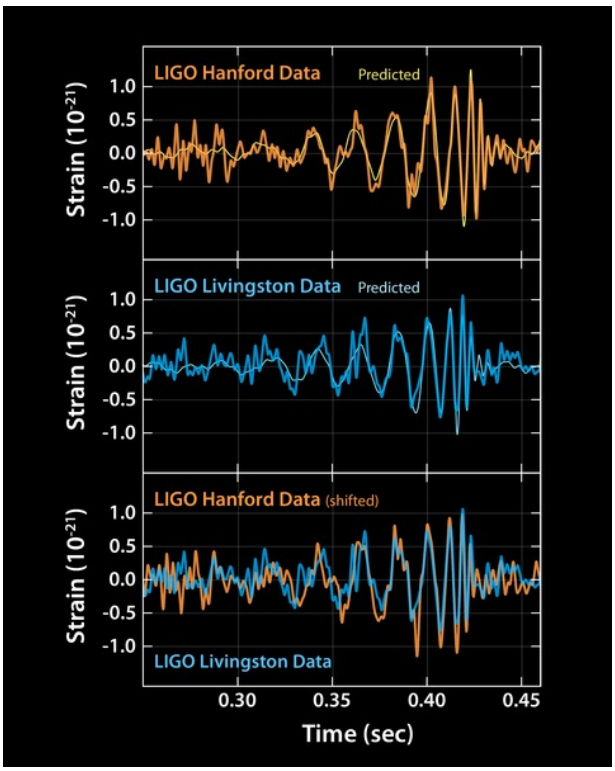

Figure 5.4: A comparison between the combined predictions of General Relativity and the sensitivities of LIGO Livingston and Hanford with the signals received at each detector from GW150914. Signal data is time-shifted to allow direct overlay of signals atop one another. Image Credit: Caltech/MIT/LIGO Lab

not so curved that a collision occurs. This yields a maximal deflection angle, α, that can be calculated for the Schwarzschild metric by [87]:

$$\alpha = \frac{4GM}{\xi c^2} \tag{5.5}$$

where ξ is the radial distance of closest approach between the photon and the Sun. This is also called the impact parameter. Newtonian mechanics also allows for the bending of light near stars in a similar manner, however, it predicts an angle half that of the General Relativity prediction [87] which allowed the Eddington experiment to directly compare its results between General Relativity and Newtonian Gravitation, yielding one of the confirmations of General Relativity.

Lens Equation

Consider the scenario of a source, S, raised from the optical plane by η being lensed by a mass, M, at an angle $\hat{\alpha}$, one may define two angles - θ which is the angle between the observer and the image I, and β which is the angle between the observer and the actual position of the source object. The magnitude of the difference between the position of the image, and the source object is given by:

$$\begin{aligned} I - \eta &= D_{OS} \tan\theta - D_{OS} \tan\beta \\ &= D \frac{\sin\alpha}{\cos\beta} \end{aligned} \tag{5.6}$$

where D is the radial distance between the source and the position where the point of closest approach is. Given that astrophysical distances are extremely large, this distance can be approximated as $D \approx D_{LS}$ through a combination of the thin lens and small angle approximations. This allows equation 5.6 to be simplified to:

$$\theta D_{OS} = \beta D_{OS} + \hat{\alpha} D_{LS} \tag{5.7}$$

This allows the definition of the lens equation as [88]:

$$\theta = \beta - \alpha \tag{5.8}$$

where α is called the reduced deflection angle and is given by $\hat{\alpha}\frac{D_{LS}}{D_{OS}}$. The lens equation thus describes the distance between the observed and actual positions of the source and holds true only so long as the distances between the observer, source, and lens are very large compared to the sizes of these objects.

Magnification Factor

Lensing is a process where light is simply deflected from one path on to another. It is a process neither of absorption or emission and that means that the intensity of the light will be conserved by the process. Additionally, considering that the distances between the lens and the observer and source are relatively constant in the system this means that the surface brightness is conserved by the source. However, what is allowed to change is the solid angle - lensing is a magnification process. The flux received from the object is a product of the surface brightness and the solid angle of the radiating object. This means that comparing the fluxes of unlensed and lensed scenarios produces a simple relation describing the amount of magnification in a single factor called the Magnification Factor, μ [89]:

$$\mu \equiv \frac{\Delta\omega}{(\Delta\omega)_0} \tag{5.9}$$

where $\Delta\omega$ is the lensed surface brightness and $(\Delta\omega)_0$ is the unlensed surface brightness.

Point Mass Lens and Einstein Angle

The simplest kind of lensing object is simply that which can be modelled by a mass concentrated in a single point. In this scenario, the impact parameter is given by $\xi = \theta D_{OL}$.

Combining this with Equation 5.5, yields a reduced deflection angle of:

$$\alpha = \frac{D_{LS}}{D_{OS}D_{OL}}\frac{4GM}{\theta c^2} \tag{5.10}$$

The point lens has an additional feature of interest. In the scenario where the source object is on the same plane as the lensing object, i.e. $\beta = 0$, the image will form a ring called an Einstein Ring. The angular radius of this ring is called the Einstein angle, θ_E. For the point mass lens, this is given by [88]:

$$\theta_E = \sqrt{\frac{D_{LS}}{D_{OS}D_{OL}}\frac{4GM}{c^2}} \tag{5.11}$$

In the more general case, where β is non-zero lensing will yield a pair of images at angular positions given by [90]:

$$\theta_\pm = \frac{\beta}{2} \pm \theta_E\sqrt{1 + \frac{\beta^2}{4\theta_E}} \tag{5.12}$$

5.2.2 Types of Lensing

Gravitational lensing occurs in the presence of any mass to varying extents as noted by the presence of the mass and distance of the lensing objects in Equations 5.11 and 5.12. The strength of the lensing can be broadly split into three categories of lensing, each with its own set of likely sources and properties. These categories are strong lensing, weak lensing, and microlensing [91].

Strong gravitational lensing occurs in the presence of large highly massive, extended sources. It is most typically the result of galaxies or galaxy clusters [91]. As might be expected by the titling of strong, the effects that it causes are the most dramatic and most directly noticeable. The signatures of strong lensing are large distortions of the image into multiple images which are spread out over the Einstein ring as defined in the previous section. The figure below, shows an example of this:

In this figure, a blue-coloured galaxy is being lensed by a bright red galaxy. The lensing is sufficiently strong that the original appearance of the galaxy is being distorted into two separate images smeared out over almost the entire Einstein Ring resulting in a horseshoe shape that is a signature of strong lensing.

Weak gravitational lensing is the regime of lensing in which the magnification factor is very close to 1, i.e. it is a very small effect meaning that multiple images are not produced. It is often the result of the combination of many small effects from foreground sources. These distortions are not identified individually, but are detected as the result of statistical analysis of background objects. This will reveal the small distortions that are the signatures of weakly lensing foreground objects. [88]

Figure 5.5: An example of strong gravitational lensing captured by the Hubble Space Telescope's Wide Field Camera 3. In this case a luminous red galaxy is distorting a bluer galaxy into two images that combine to form a 'horseshoe' shape. Image Credit: ESA/Hubble & NASA

Gravitational microlensing is the final category of gravitational lensing. It is the result of smaller compact objects such as stellar mass black holes or stars. The scale of mass of these objects results in multiple images with angular separations on the scales of milli-arcseconds which means that instead of being resolvable as separate images as is the case in strong lensing, the images are only resolvable as one object. Instead the change that is observable is one of apparent brightness. [88]

5.2.3 Lensing Profiles

Thus far, the only lensing object that has been considered is that of the point mass. However, real world objects are not singular points - they are extended objects with densities that may be able to vary throughout. This is contained within the density profile, $\rho(\mathbf{x})$. Several density profiles exist to describe various scenarios of extended object. However, in most scenarios, one of the approximations that is made is called the thin-lens approximation in which the thickness of the lens is ignored due to the assumption that this thickness is negligible compared to the distances between the source and the lens, and the lens and the observer. This approximation means that as opposed to the complete three dimensional density function, the surface mass density, $\Sigma(\boldsymbol{\xi})$, is considered instead.

The surface mass density is defined as the projection of the full density onto a two dimensional plane at the centre of mass of the lensing object. When the mass distribution is spherically symmetric, the use of this simplifies calculations due to the surface mass density depending only upon the modulus of the impact parameter. This allows the lens equation to be rewritten for the case of an extended mass distribution as:

$$\beta = \theta - \frac{M(\theta)}{\pi D_{OL}^2 \theta \Sigma_{cr}} \tag{5.13}$$

where Σ_{cr} is called the critical surface mass density and is defined as [88] [91]:

$$\Sigma_{cr} = \frac{c^2 D_{OS}}{4\pi G D_{OL} D_{OS}} \tag{5.14}$$

The final thing to be calculated from each of these profiles that is necessary is the reduced mass. This is the amount of mass that is enclosed inside of a circle of radius ξ in the projected plane. It is defined thusly :

$$M(\xi) \equiv 2\pi \int_0^{\xi} \xi' \Sigma(\xi') d\xi' \tag{5.15}$$

Singular Isothermal Sphere (SIS)

Galaxies are most regularly modelled as large, extended objects containing luminous matter surrounded by a dark matter halo. The evidence for these halos is the flat rotation curve seen in galaxies. To recreate this rotation curve in a symmetric mass distribution, there are several possible density profiles. However, the most simple and the most widely used is the Singular Isothermal Sphere (SIS) profile which is given by [92]:

$$\rho_{SIS}(r) = \frac{\sigma^2}{2\pi G r^2} \tag{5.16}$$

Here σ is the one-dimensional velocity distribution of stars around the galaxy or cluster being modelled i.e. the velocity dispersion. Whilst useful for the recreation of the rotation curve, the SIS does suffer from an nonphysical property in the form of a singularity at the centre of the distribution ($r = 0$). Using this density profile, yields a surface mass density profile of:

$$\Sigma_{SIS}(\xi) = \frac{\sigma^2}{2G\xi} \tag{5.17}$$

The reduced mass, then follows easily and is calculated analytically as:

$$M_{SIS}(\xi) = \frac{\pi \sigma^2}{G} \xi \tag{5.18}$$

Navarro, Frenk, and White (NFW)

As is outlined in §3.2.6, the most common cosmological models including the Standard Model of Cosmology take the approach that dark matter is cold. They also take the approach that dark matter is collisionless. The work of Navarro, Frenk, and White indicated that dark matter halos in this model could be described by scaling a 'universal' profile. The NFW profile is given by [92]:

$$\rho_{NFW}(r) = \frac{r_s}{r/r_s(1+r/r_s)^2} \tag{5.19}$$

where ρ_s is the central density and r_s is the characteristic scale radius of the profile. As can be seen, once again the profile has a singularity in the centre.

The mass profile leads to the surface mass density profile of [92]:

$$\Sigma_{NFW}(\xi) = \begin{cases} \frac{2r_s\rho_s}{x^2-1}\left[1 - \frac{2}{\sqrt{1-x^2}}\ \text{arctanh}\ \sqrt{\frac{1-x}{1+x}}\right] & \text{if } x < 1 \\ \frac{2r_s\rho_s}{3} & \text{if } x = 1 \\ \frac{2r_s\rho_s}{x^2-1}\left[1 - \frac{2}{\sqrt{1-x^2}}\arctan\sqrt{\frac{x^2-1}{1+x}}\right] & \text{if } x > 1 \end{cases} \tag{5.20}$$

where for spherical symmetry, $x = \xi/r_s$. This leads to a reduced mass of:

$$M_{NFW}(\xi) = 4\pi r_s\rho_s \begin{cases} \left[\frac{2}{\sqrt{1-x^2}}\ \text{arctanh}\sqrt{\frac{1-x}{1+x}} + \ln(\frac{x}{2})\right] & \text{if } x < 1 \\ \left[1 + \ln(\frac{1}{2})\right] & \text{if } x = 1 \\ \left[\frac{2}{\sqrt{x^2-1}}\arctan\sqrt{\frac{x-1}{1+x}} + \ln(\frac{x}{2})\right] & \text{if } x > 1 \end{cases} \tag{5.21}$$

Burkert Profile

As has been noted for both the Singular Isothermal Sphere and the Navarro, Frenk, White profiles despite their wide use and simplicity, they are both unphysical due to central singularities. One solution to this is to create a profile with a finite density core. One profile that does this is called the Burkert profile [93]:

$$\rho_B(r) = \frac{\rho_s}{(1+r/r_s)(1+(r/r_s)^2)} \tag{5.22}$$

where in a similar fashion to the NFW profile, r_s and ρ_s represent the characteristic scale radius and central density for the dark matter profile in consideration. The Burkert profile is useful for the fact that at large radii it agrees extremely well with the NFW profile.

Soliton Profile for Scalar Field Dark Matter

As was discussed in §3.2.6, one of the possible models of dark matter is viewing it as a scalar field. In this scenario, there are additional density profiles that may be considered. Through intensive computational analysis of galaxy cores, a fitted analytical profile called the soliton was constructed and is given as follows [94]:

$$\rho_{sol}(r) = \frac{\rho_s}{[1 + (r/r_s)^2]^8} \tag{5.23}$$

where once again, r_s is the characteristic scale radius and ρ_s is the central density. This gives the core, but to create a more complete density profile it can be given an NFW tail. This means that the complete profile is given by [94]:

$$\rho(r) = \Theta(r_\epsilon - r)\rho_{sol}(r) + \Theta(r - r_\epsilon)\rho_{NFW}(r) \tag{5.24}$$

where Θ is a step function and r_ϵ is the matching radius which is the radius at which the transition between the two profiles occurs. To ensure that the profile is continuous, the transition radius is defined such that [94]:

$$\rho_{sol}(r_\epsilon) = \rho_{NFW}(r_\epsilon) = \epsilon\rho_s \tag{5.25}$$

The surface mass density profile for the more complete profile is calculated by performing [94]:

$$\Sigma(\xi) = 2\int_0^\infty \left[\Theta(r_\epsilon - r)\rho_{sol}(r) + \Theta(r - r_\epsilon)\rho_{NFW}(r)\right] dz \tag{5.26}$$

where the integral is done over the line of sight z and encloses a circle of radius ξ_* which necessitated a change of variables to $r = \sqrt{\xi^2 + z^2}$. Splitting this integral to its constituent parts yields a normalised surface mass density of [94]:

$$\Sigma_*(\xi_*, r_{\epsilon*}, \alpha_{NFW}) = 2\begin{cases} \int_0^{\sqrt{r_{\epsilon*}^2 - \xi_*^2}} \frac{dz}{(1+r_*^2)^8} + \frac{r_{\epsilon*}(1+\alpha_{NFW}r_{\epsilon*})^2}{(1+r_{\epsilon*}^2)^8} \int_{\sqrt{r_{\epsilon*}^2 - \xi_*^2}}^\infty \frac{dz_*}{r_*(1+\alpha_{NFW}r_*)^2} & \text{if } \xi_* < r_{\epsilon*} \\ \frac{r_{\epsilon*}(1+\alpha_{NFW}r_{\epsilon*})^2}{(1+r_{\epsilon*}^2)^8} \int_0^\infty \frac{dz_*}{r_*(1+\alpha_{NFW}r_*)^2} & \text{if } \xi_* > r_{\epsilon*} \end{cases} \tag{5.27}$$

In the case of the scalar field dark matter scenario, the reduced mass is not used. Instead the quantity is called the projected mass which is the mass enclosed in a cylinder with the radius of the Einstein ring over the line of sight projected onto the lens plane. This is given by [1]:

$$m(\xi_*) = 4\pi \begin{cases} \frac{1}{14}\left[B(\arctan r_{\epsilon*}) - \frac{B(\arctan q1)}{(1+\xi_*^2)^{13/2}} + \frac{\sqrt{r_{\epsilon*}^2 - \xi_{\epsilon*}^2} - r_\epsilon}{(1+r_\epsilon^2)^7}\right] + \\ \frac{\rho_{NFW*}}{\alpha_{NFW}^3}\left[\ln\left(\frac{y+\sqrt{y^2-x^2}}{2(y+1)}\right) + \frac{y-\sqrt{y^2-x^2}}{y+1} + f(x,y)\right] & \text{if } \xi_* < r_{\epsilon*} \\ \frac{1}{14}\left[B(\arctan r_{\epsilon*}) - \frac{r_\epsilon}{(1+r_\epsilon)^7}\right] + \frac{\rho_{NFW*}}{\alpha_{NFW}^3}\left[\ln\left(\frac{1}{y+1}\right) + \frac{y}{y+1} + m_{NFW}(x)\right] & \text{if } \xi_* > r_{\epsilon*} \end{cases} \tag{5.28}$$

The symbol definitions are given in the following Table, and B is a function defined by [1]:

$$\begin{aligned} B(u) &\equiv \int_0^u \cos^{12} u' du' \\ &= \frac{1}{122880}[27720u + 23760\sin(2u) + 7425\sin(4u) + 2200\sin(6u) + 495\sin(8u) \\ &+ 72\sin(10u) + 5\sin(12u)] \end{aligned} \tag{5.29}$$

Parameter	Meaning
x	$\alpha_{NFW}\xi_*$
y	$\alpha_{NFW}r_{\epsilon*}$
$q1$	$\sqrt{(r_{\epsilon*}^2 - \xi_*^2)/(1+\xi_*^2)}$
m_{NFW}	Equation 5.21

Table 5.1: The necessary additional definitions for Equation 5.28

5.2.4 Lensing of Gravitational Waves

Gravitational lensing does not only apply to light, it also applies to gravitational waves. In this case, the change in the signal is defined not by a magnification factor which defines a change in brightness, but instead by the change in wave amplitude which is leads to the definition of an amplification factor, termed $F(w, \boldsymbol{\eta})$ as follows:

$$F(w, \boldsymbol{\eta}) = \frac{\phi_{obs}^L(w, \boldsymbol{\eta})}{\phi_{obs}(w, \boldsymbol{\eta})} \tag{5.30}$$

where the numerator and denominator are the lensed and unlensed wave amplitudes respectively and $\boldsymbol{\eta}$ and w are the source position and the dimensionless frequency respectively. This leads to a very simple relation between the lensed and unlensed strains [4]:

$$h^L(f) = h(f) \times F(f) \tag{5.31}$$

This simple relation is possible due to the uncorrelated nature between the parameters defining the signal and the parameters defining the amplification factor with the exception of the frequency. It is also this simple expression which shows the value of gravitational wave lensing as a probe of dark matter. Comparing any lensed signals with the expected unlensed signal would yield the amplification factor, which is dependent upon the lensing profile.

The individual amplification factors for each lensing profile are calculated from one more general expression. This is simplest for the case of axially-symmetric lenses where it is given by [95]:

$$F(w,y) = -iwe^{iwy^2/2}\int_0^\infty dx x J_0(wxy) \times \exp\left[iw\left(\frac{1}{2}x^2 - \psi(x) + \phi_m(y)\right)\right] \tag{5.32}$$

where $\psi(x)$ is a function called the lensing potential, $\phi_m(y)$ is the time delay of the shortest possible path to the observer, and the dimensionless frequency is defined by [96]:

$$w = \frac{D_{OS}}{D_{LS}D_{OL}}\xi_0^2(1+z_l)\omega \tag{5.33}$$

where ξ_0 is a normalisation constant.

Point Mass Lens

The Point Mass Lens is the simplest possible case for the calculation of the amplification factor. It is in fact, the only lensing profile, where Equation 5.32 can be calculated analytically. In this case, the normalisation constant is chosen to correspond to the Einstein radius and is given by [96]:

$$\xi_0 = \left(\frac{4M_L D_{OL} D_{LS}}{D_{OS}}\right)^{1/2} \tag{5.34}$$

This leads to a dimensionless frequency given by $w = 4M_{Lz}\omega$ where M_{lz} is the redshifted lens mass. This yields a solution for the amplification factor as [96]:

$$F(w,y) = \exp\left[\frac{\pi w}{4} + i\frac{w}{2}\left\{\ln\left(\frac{w}{2}\right) - 2\phi_m(y)\right\}\right]\Gamma\left(1-\frac{i}{2}w\right){}_1F_1\left(\frac{i}{2}w;1;\frac{i}{2}wy^2\right) \tag{5.35}$$

where ${}_1F_1$ is the confluent hypergeometric function of the first kind and $\phi_m(y)$ is given by $(x_m - y)^2/2 - \ln x_m$ where $x_m = (y + \sqrt{y^2+4})/2$.

Singular Isothermal Sphere

As was mentioned previously, the Point Mass case is the only case where the amplification factor integration can be solved analytically. For the SIS case, along with the rest of the following cases, the integration must be solved numerically. In this case, the lensing potential is given by $\psi(\mathbf{x}) = |x|$, and the shortest time delay is given by $\phi_m(y) = y + \frac{1}{2}$. The normalising factor, ξ_0 is given by $4\pi v^2 D_{OL} D_{LS}/D_{OS}$ which means that in order to yield the same definition of the dimensionless frequency, the mass M_{lz} for the SIS is given by $4\pi^2 v^4 (1+z_l) D_{OL} D_{LS}/D_{OS}$. [96]

Navarro, Frenk, and White

The NFW lens, similarly to the SIS case must have the amplification factor calculated numerically. However, it introduces an additional complication as compared to the the SIS, because the time delay constant must also be numerically calculated because the lens equation cannot be solved analytically. The lensing potential is also more complex, being given by [97]:

$$\psi(x) = \frac{k_s}{2} \begin{cases} \left[\left(\ln \frac{x}{2}\right)^2 - 4 \left(\operatorname{arctanh} \sqrt{\frac{1-x}{1+x}}\right)^2 \right] & \text{if } x \leq 1 \\ \left[\left(\ln \frac{x}{2}\right)^2 - 4 \left(\arctan \sqrt{\frac{x-1}{1+x}}\right)^2 \right] & \text{if } x \geq 1 \end{cases} \tag{5.36}$$

where the parameter k_s is defined as:

$$k_s = 16\pi \rho_s \left(\frac{D_{OL} D_{LS}}{D_{OS}} \right) r_s \tag{5.37}$$

An additional difference between the NFW and SIS cases is that in the NFW case, instead of choosing the normalising constant to be the Einstein radius, the characteristic radius of the profile is chosen instead, i.e. $\xi_0 = r_s$

Soliton

As was mentioned in the previous section, there are two profiles that can be generated based around the Soliton model for Scalar Field Dark Matter: the soliton core alone, and the more complete Soliton core + NFW tail. The calculations for both must be performed numerically, however, the calculations for the complete profile are extremely complex and thus only the core alone is considered.

In this case, the lensing potential is calculated by evaluation of [1]:

$$\psi(x) = 2 \int_0^x x' dx' \kappa(x') \ln \left(\frac{x}{x'} \right) \tag{5.38}$$

where $\kappa(x) = \Sigma(x)/\Sigma_{crit}$ and: [1]

$$\Sigma(x) = \rho_s r_s \frac{429\pi}{2048}(1+x^2)^{-15/2} \tag{5.39}$$

and $x = \xi/r_s$. Ultimately this yields a solution to Equation 5.38 of [1]:

$$\psi(x) = 2k_s \frac{429\pi}{2048}\left[\frac{1}{13}\left(\ln\left(\sqrt{x^2+1}+1\right) - \sum_{n=0}^{5}\frac{1}{(2n+1)(1+x^2)^{n+1/2}}\right) + \frac{6508 - 3465\ln(2)}{45045}\right] \tag{5.40}$$

where $k_s = \rho_s r_s/\Sigma_{crit}$.

Chapter 6

Code Development

The project largely concerned the development of python based software that would be able to determine from gravitational wave signal data whether that signal was lensed, and if it was lensed what kind of amplification factor was the most likely. This would allow gravitational wave detectors to become an excellent probe of dark matter as outlined in the previous chapter. The foundations for the code were a series of amplification factor calculation codes written in Mathematica by Antonio Herrera Martin [1]. The following concerns how the python code was developed and tested for accuracy.

6.1 Code Overview

6.1.1 Toolkits and Language Used

The primary programming language in which the code was written was python. This maximises a combination of being widely available, efficient, and able to perform the necessary tasks. However, due to the extreme oscillatory nature of some of the integrals outlined in §5.2.4, the numerical quadrature algorithms widely available in python were unable to perform sufficiently accurately for use - necessitating the continued use of Mathematica to perform these integrations.

For the purposes of the generation and analysis of gravitational wave signals, the LSC Algorithm Library Suite (LALSuite) [3] implementation in python is an extremely effective tool - however, it can prove to be difficult for newcomers to learn and use effectively. To more easily generate these signals, as well as easily perform the data analysis using nested sampling in a fully integrated manner, bilby was used [2]. Bilby describes itself as "a user-friendly Bayesian inference library" [2] and does have a direct focus on gravitational waves by wrapping the power of the LALSuite into easy to use signal generation functions, as well as being able to access real signal data simply.

6.1.2 Code Procedure

The code is constructed in a modular fashion - allowing various different start and end states to allow the production of results at any stage of analysis. The first step taken in the code after the setup preamble, is the definition of the injection parameters. These are the model parameters that are used to define the priors for any of the parameters not being estimated, and function as the 'true' parameters when the data is being simulated.

After this step, is the generation or loading in of amplification factor grids for any of the lensed models being tested. These grids are the result of calculating the amplification factors for varying values of dimensionless frequency and impact parameter. These grids are then separated into their real and imaginary components allowing the creation of interpolation functions which can be combined into a total interpolation function. Whilst it would be more complete to calculate the amplification factor for as many points as are generated for the gravitational wave signal, the calculations are computationally intensive and it is much more time-efficient for many runs to have a pre-generated grid that can be used.

With the lensing amplification factors available, the next stage is to generate the gravitational wave signals using bilby. Bilby allows for the use of many common waveform models such as 'IMRPhenomD' or 'TaylorF2'. These initial generators can then also be modified allowing native compatibility with the lensed models, by simply defining them as the product of the strain and the amplification factors. Bilby then allows for the use of the sensitivity curves of many gravitational wave detectors - both operational and planned to include reasonable noise levels and thus generate signals that look realistic to the detectors that would see them.

The following stages are the preparation for the nested sampling procedures, the generation of priors and likelihoods. The priors are contained in a single dictionary which is able to contain ranges in different formats, primarily Uniform priors were chosen for those priors that are of interest to be analysed. For those priors that are not of interest to be analysed, the prior is given simply as the value chosen in the injection parameters in the first step. The likelihoods can be selected from a set of predefined functions. For all tests, the function used was 'GravitationalWaveTransient' using the waveform generators and detectors chosen.

With each of these defined, the nested sampling algorithms can then be used. Bilby allows for the picking of several kinds of nested sampling implementations that are native to python. The primary sampler used in testing was the default 'dynesty'. This yields both the samples, and a complete corner plot for all of the of the tested parameters for each of the models tested. The samples contain each of the parameters tested at that sample, as well as most crucially the evidence for that sample which can then be compared with the evidences for the other models to determine which of the models is the most likely.

6.2 Code Testing

6.2.1 Reproduction of Amplification Factor Generation

The first testing procedure was to replicate the results achieved by the work of Antonio Herrera Martin in generating the amplification factors. The initial intent had been to completely translate the code from the original Mathematica notebooks into python. The purpose of this translation was to enable a greater automation potential than the notebooks - as these need to be run in the individual segments. An additional benefit of this would be that python is freely available where Mathematica is a proprietary language owned and developed by Wolfram Research.

Being the simplest of the amplification factors and the only one which can be calculated completely analytically, the point mass could be successfully replicated in python. The visual comparison between Figures 6.1 and 6.2 suggests agreement, as can be seen in the presence of similar features such as the $y = 0.1$ graphing containing 3 complete peaks in the given range. For a more comprehensive test between the two, the Table below shows a comparison between the results of the amplification factor calculation for a sample of points covering a range of dimensionless frequency and impact parameter. Using the data from that table yields an average difference between the Mathematica and the python values of $-7.7 \times 10^{-15} - 1.8 \times 10^{-14}i$ which is a magnitude of 1.9×10^{-14}. Given that the order of the magnitudes of the amplification factor are about unity this is a negligible disagreement and justifies the python code as a complete replacement for the Mathematica.

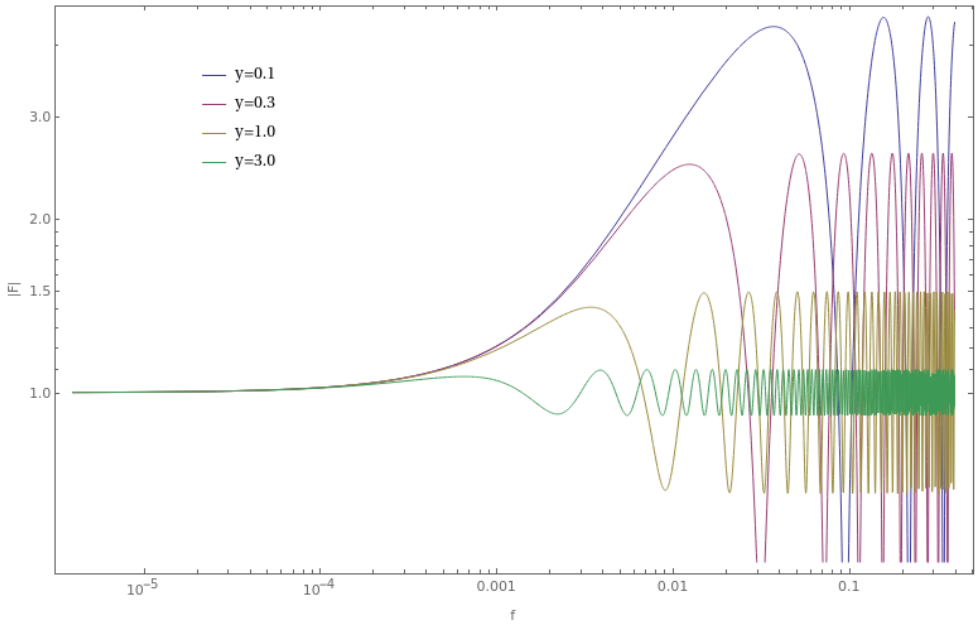

Figure 6.1: Point Mass Lens Amplification Factor generated by Antonio Herrera Martin's original Mathematica notebook for varying impact parameters using a lens mass of $10M_{\odot}$

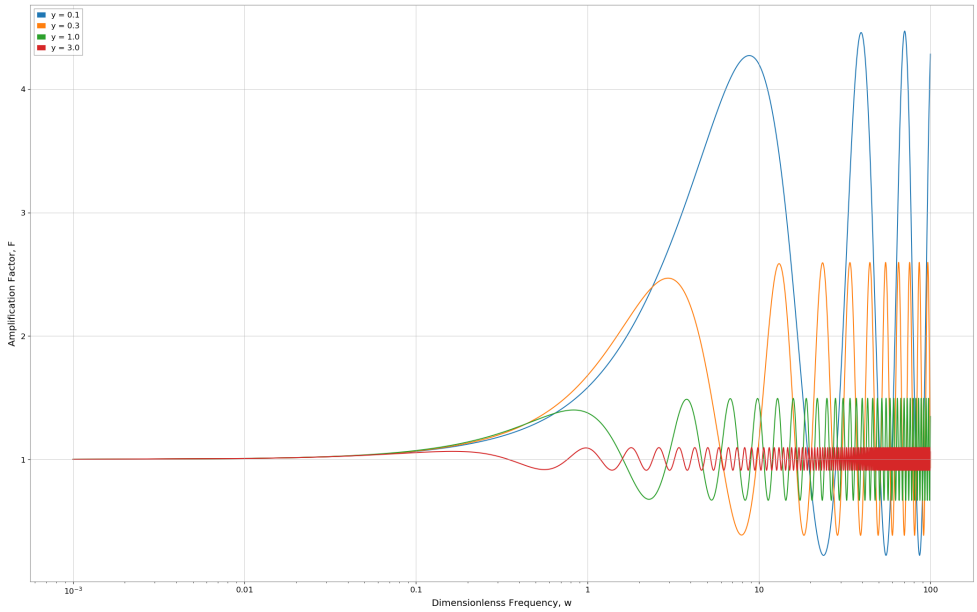

Figure 6.2: Point Mass Lens Amplification Factor generated by python code in the same configuration as for the previous figure

w	y	Difference Between Mathematica Value and Python Value
0.1	0.1	$1.33227 \times 10^{-15} - 1.11022 \times 10^{-16} i$
0.1	0.3	$1.11022 \times 10^{-15} - 8.32667 \times 10^{-17} i$
0.1	1.0	$1.11022 \times 10^{-15} + 1.38778 \times 10^{-17} i$
0.1	3.0	$8.88178 \times 10^{-16} + 6.59195 \times 10^{-17} i$
10	0.1	$1.24345 \times 10^{-14} - 2.99760 \times 10^{-15} i$
10	0.3	$1.77636 \times 10^{-15} - 4.44089 \times 10^{-15} i$
10	1.0	$4.66294 \times 10^{-15} - 1.27676 \times 10^{-15} i$
10	3.0	$3.33067 \times 10^{-15} - 1.05471 \times 10^{-15} i$
100	0.1	$-7.90479 \times 10^{-14} - 2.07501 \times 10^{-13} i$
100	0.3	$-1.37668 \times 10^{-14} - 8.21565 \times 10^{-15} i$
100	1.0	$-1.46549 \times 10^{-14} + 7.93809 \times 10^{-15} i$
100	3.0	$-1.19904 \times 10^{-14} + 4.27436 \times 10^{-15} i$

Table 6.1: Table showing the difference between the Mathematica and python generated values for the point mass lens. The extremely minimal difference indicates that the python is able to successfully replicate the Mathematica's results

Unfortunately, the point mass model was the only one in which the python was able to successfully replicate the Mathematica. The reason for this is the highly oscillatory nature of the integrals that must be numerical computed using quadrature algorithms. To perform quadrature on such functions requires more sophisticated and dedicated quadrature

methods that were unavailable in any checked python package. As a result, in order to retain the more automated benefits brought by python - the calculation commands in the Mathematica notebooks were converted into simple Wolfram Scripts which allow them to be run in the terminal, and the running and handling of the results of these scripts can be done within python. This allows python to use the Wolfram Kernel as a backend for the construction of amplification factor grids and seamlessly to the user integrate this data into the analysis processes. Whilst this achieves the necessary goal of improving the automation, it does limit the generation of the more complex models to those with access to Wolfram licenses.

For the SIS case, which is the most simple of the three more complex cases, the figures (6.3 and 6.4) below show the results of the Mathematica and python attempts to calculate the amplification factors. Similarly to the point mass case, in a visual comparison it is difficult to notice many distinct differences between the two. However, comparison of the individual values yield much more significant differences between the two than is present for the point mass case. The table below shows the average difference between the python and Mathematica values for the magnitude of the amplification factor for three values of impact parameter, the figure then shows the full comparisons between the two.

Given the scale of the SIS amplification factor ranges between order unity and order 10, and given the need for stringent precision when it comes to gravitational wave detection, the average differences shown in Table 6.2, and shown visually in full in Figure 6.5 are too large to be ignored. As a result of this, the Mathematica calculation code remained and was converted into script form to be used as a backend for result generation in the greater python infrastructure.

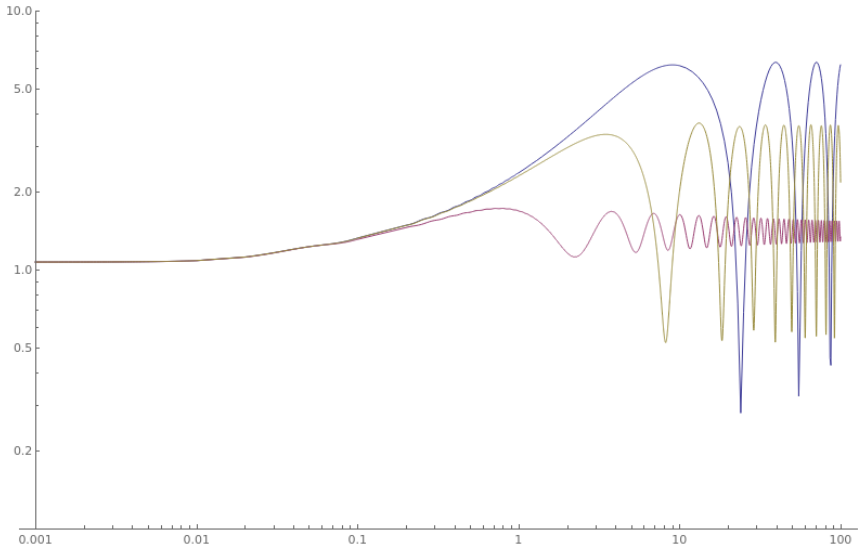

Figure 6.3: Mathematica calculation of the SIS Amplification Factor for y=0.1 (blue), y=0.3(green), and y=1.0(pink) resepctively

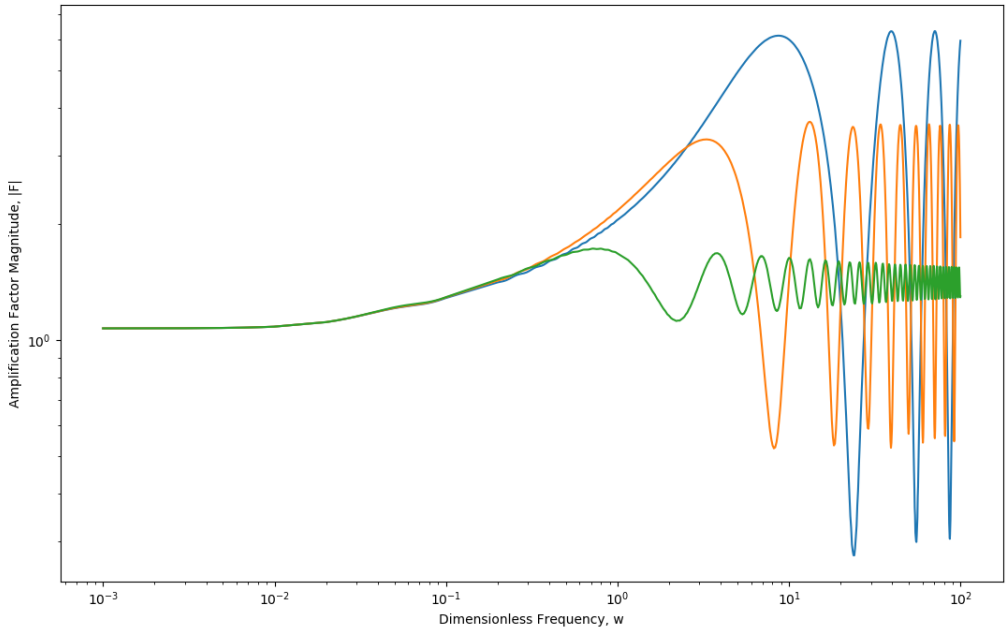

Figure 6.4: Python calculation of the SIS Amplification Factor for the same set of impact parameter values

Impact Parameter, y	Mean Difference Between Python and Mathematica Values
0.1	-0.402585
0.3	0.105699
1.0	-0.051046

Table 6.2: Values of Mean Difference Between python and Mathematica calculated values of the magnitude of the amplification factor for different impact parameters for the SIS case

The Navarro, Frenk, and White Profile alongside the Soliton profile are more computationally complex in terms of the integration that needs to be performed, and as a result of the failure to fully pythonise the SIS case, these were not converted and simply given scripting form. For the NFW case, this proved a simple task as whilst it is more complex to compute, it is not significantly different to the SIS case in terms of how these computations are performed. In both of these cases, the computations are inexpensive in time enough that if desired, one can generate the values needed on a per run basis. Figure 6.6 below shows the amplification factor plot for the NFW model.

The final and most complex profile, is the Soliton profile. Unlike the NFW profile, the Soliton is so computationally expensive to run that it adds a great deal of computational expense to create the data on a per-run basis. Instead, a single run can be performed to generate a very large data table to vastly increase efficiency in performing analysis. The

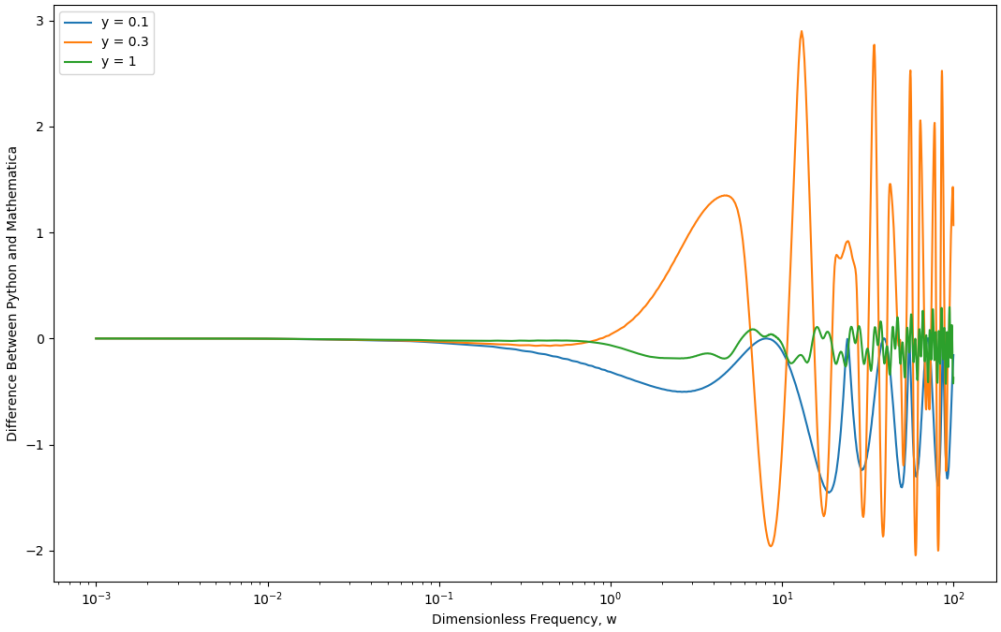

Figure 6.5: Value of the Difference Between python and Mathematica calcuated values of the magnitude of the amplification factor for different impact parameters for the SIS case

computational expense also requires that the calculation be broken up into many smaller chunks and then combined together. In a similar fashion as to the other models, Figure 6.7 below shows the amplification factor calculated for three different impact parameter values.

In the case of each of these profiles, the calculation over a grid of frequency and impact parameter values can then be converted into a two dimensional interpolating function which increases the efficiency of using the data points should any additional frequencies be needed.

6.2.2 Toy Model Problem

Given the major importance of the evidence in performing model selection using the nested sampling method outlined in Chapter 4, it is important to be able to trust that the evidence calculated by the Bilby samplers is in fact correct. For the vast majority of cases, the evidence must be calculated numerically making this a slightly more difficult endeavour. However, there are some simple cases where the evidence can be calculated analytically. This allows the trustworthiness of the evidence results to be evaluated.

The toy problem that was analysed in this case was adapted from an exam question given by the University of Glasgow concerning the value of a parameter z based upon observation. A first observation yields that the parameter has an estimated value z_1 with

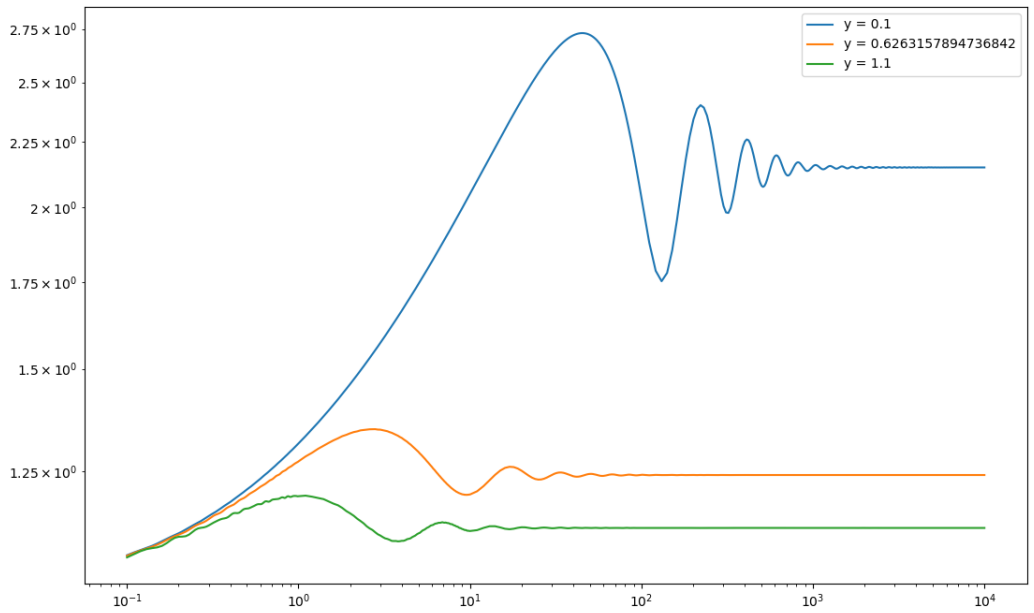

Figure 6.6: Amplification Factor for the Navarro, Frenk, and White profile calculated at three values of the impact parameter, using a ks value of 1, and normalising to the critical value of y.

a standard deviation of σ. Assuming that the uncertainty in the data is Gaussian, then the posterior will be given by:

$$p(z|z_1) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(z-z_1)^2}{2\sigma^2}\right) \tag{6.1}$$

If a second measurement is taken yielding a value of z_2 with the same standard deviation of σ then the posterior becomes:

$$p(z|z_1, z_2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(z-z_1)^2 + (z-z_2)^2}{2\sigma^2}\right) \tag{6.2}$$

In this latter case, the evidence, E, becomes:

$$E \propto \exp\left(\frac{-\left(\frac{1}{2}(z_1^2 + z_2^2) - z_1 z_2\right)}{2\sigma^2}\right) \tag{6.3}$$

where the constant of proportionality is given by the value of the evidence that results from $z_1 = z_2$ i.e. in which the term in the expression above would be 0.

The procedure to test this in bilby to analyse the result was to set up the model as described above. Firstly the code would be run ten times for the scenario of the two values being equal. This allows for the constant of proportionality to be established with

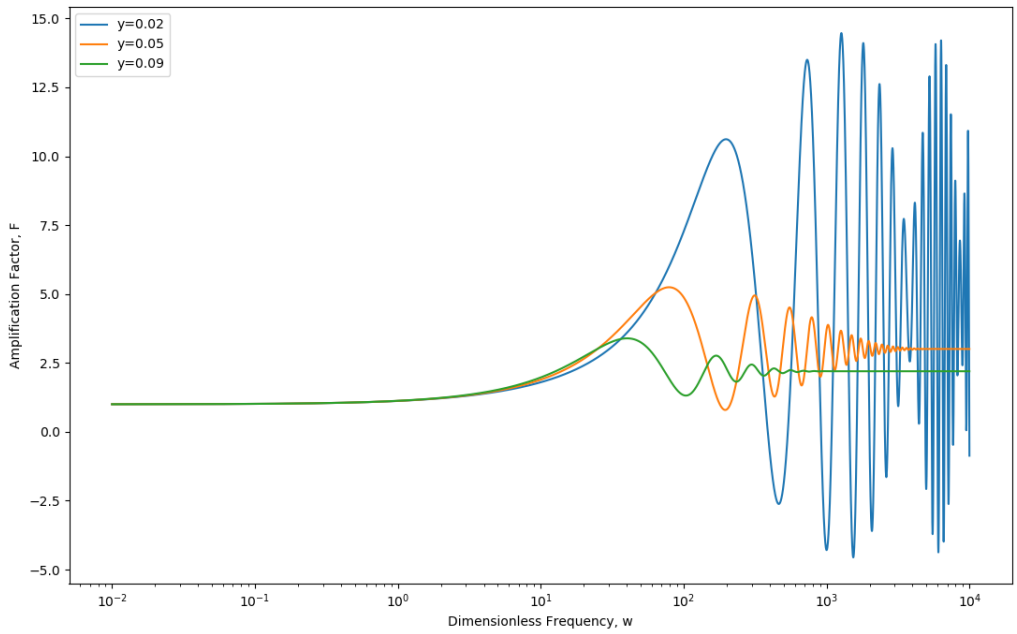

Figure 6.7: Amplification Factor for the Soliton lens model at three different values of impact parameter

reasonable accuracy. The code could then be run for more general cases and the resulting evidence values could be compared with the analytically calculated values. Table 6.3 below gives the values of the analytical and bilby calculated evidence for a number of values of $|z_1 - z_2|$.

Figures 6.8, 6.9, and 6.10 show the visual representations of the table with 6.8 showing the raw values of the bilby and analytically calculated evidences, and the latter two showing the difference between these values both in absolute and fractional terms. With a maximal difference of approximately 12%. there is reasonable agreement between the values, justifying trust in the evidences generated by bilby in further use cases.

6.3 Comparison with Cao et al. (2014)

The final testing procedure was to compare the results of comparing the lensed and unlensed signals with previous work comparing lensed and unlensed signals. Specifically, the comparison was made with Cao et. al in their 2014 paper [4]. In full using the 'TaylorF2' waveform template, they compared the lensed and unlensed signals for a black hole binary, a neutron star binary, and a black hole - neutron star binary, as they would be detected by the Advanced LIGO detectors under the assumption there was no noise. They did this for lensing masses of $10M_\odot$, $1000M_\odot$ and $4.4 \times 10^6 M_\odot$ corresponding to the object being

$\lvert z_1 - z_2 \rvert$	Analytical Evidence	Bilby Evidence
0 (Single Evaluation)	0.141938	0.126478
0.1	0.139738	0.150273
0.2	0.133339	0.124978
0.3	0.123318	0.133503
0.4	0.110542	0.108239
0.5	0.096040	0.090016
0.6	0.080874	0.078248
0.7	0.066007	0.062319
0.8	0.052216	0.047024
0.9	0.040035	0.041730
1.0	0.029752	0.028302
1.1	0.021429	0.021039
1.2	0.014960	0.015296
1.3	0.010123	0.010303
1.4	0.006639	0.006605
1.5	0.004220	0.004479
1.6	0.002599	0.002902
1.7	0.001552	0.001354
1.8	0.000898	0.001013
1.9	0.000504	0.000501
2.0	0.000274	0.000267

Table 6.3: Comparison between the evidences produced by analytical calculation and bilby for the Toy Model problem

a stellar mass black hole, an intermediate mass black hole, and a supermassive black hole respectively and using the point mass and SIS lensing profiles.

In order to compare the signals quantitatively, as opposed to visually, they calculated a fitting factor between the lensed and unlensed signals. This fitting factor was based on the inner product as follows [4]:

$$FF \equiv \max_{t_c,\phi_c,\mathcal{M},\nu} \frac{(h_u(t_c,\phi_c,\mathcal{M},\nu)|h_l)}{\sqrt{(h_u|h_u)(h_l|h_l)}} \tag{6.4}$$

where h_u and h_l are the unlensed and lensed strains respectively. The closer this fitting factor is to unity, the more similar the signals are. In Cao et al [4], this was used as a guide to indicate how lensed signals might bias parameter estimation compared to unlensed signals. Table 6.4 below shows the values obtained by Cao et al in their paper, and the results obtained by using the bilby code using the 'TaylorF2' waveform and using no noise when injecting into the detector for the point mass and SIS case below for a $1000M_\odot$ lens mass. Figure 6.11 below shows the comparison between the lensed and unlensed signals used to calculate the fitting factor for the point mass lens model using $y = 0.1$ and a lens mass of $1000M_\odot$.

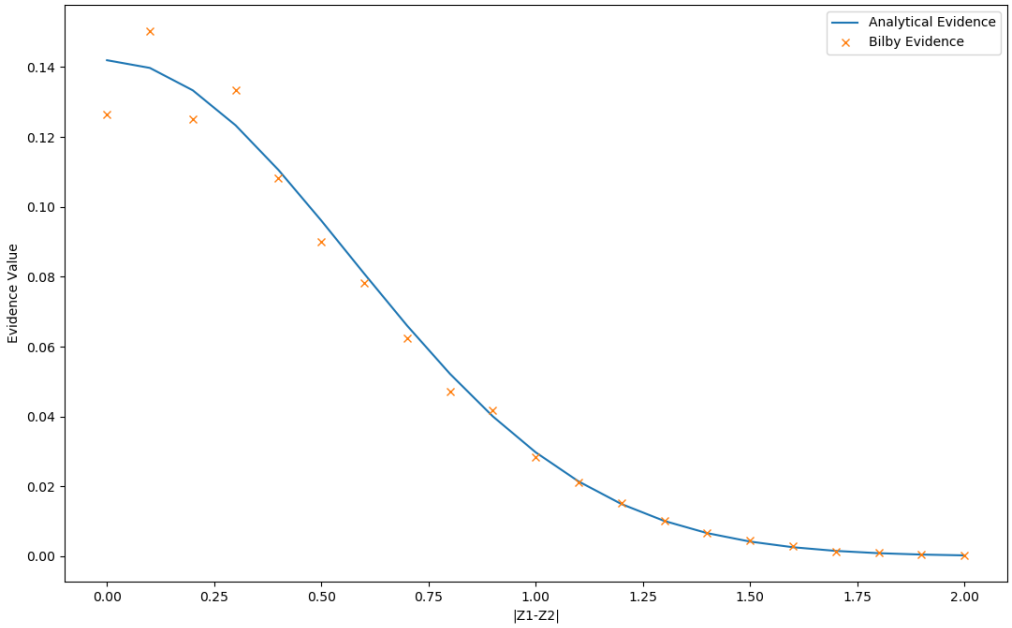

Figure 6.8: Comparison between the analytically calculated evidences (in blue) and the bilby calculated evidences (in orange) for the Toy Model problem

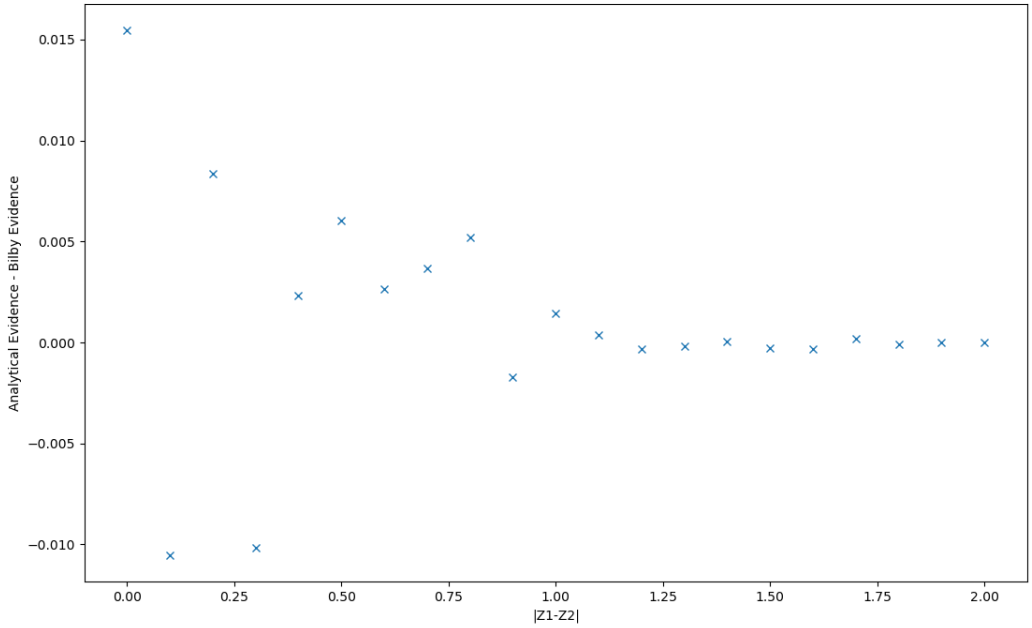

Figure 6.9: Values of the absolute difference between the evidences calculated analytically and those produced by bilby for the Toy Model problem

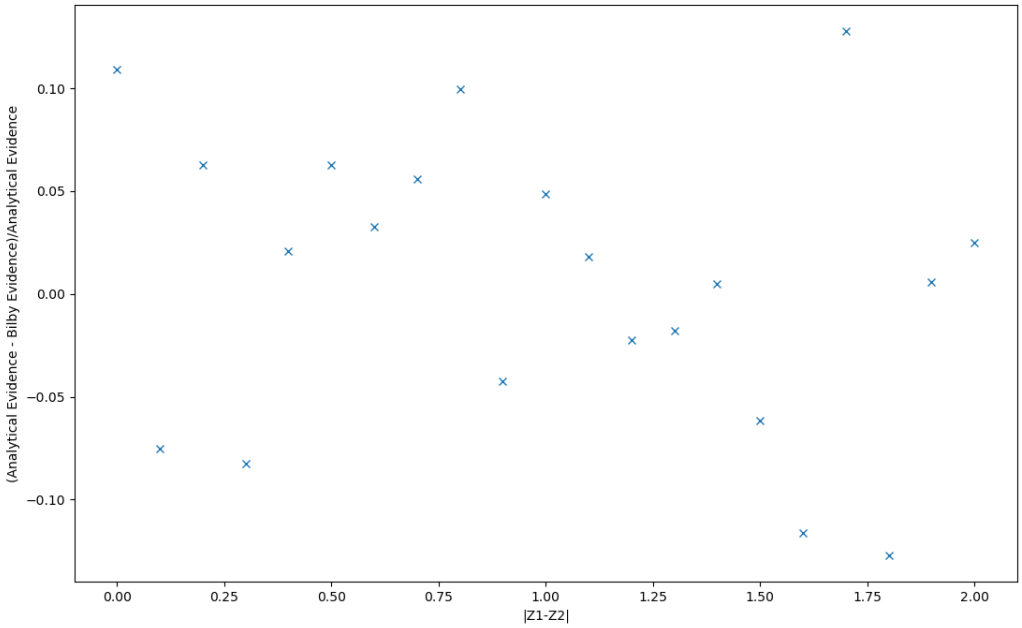

Figure 6.10: Fractional Differences between analytically calculated evidence values and bilby calculated evidence values for the Toy Model problem

The comparison between the fitting factors, results in similar values for the fitting factor above $y = 0.1$, with a larger difference at lower y values. Based on this comparison, it would justify the usage of the bilby code where y is 0.1 or above. As a result of this, the cases considered in the next chapter where the code is used, all simulate cases where the impact parameter is above this lower boundary.

Model	FF (Cao)	FF (Bilby)
Point Mass $y = 0.01$	0.954	0.757
Point Mass $y = 0.03$	0.973	0.852
Point Mass $y = 0.1$	0.832	0.895
Point Mass $y = 0.3$	0.792	0.800
SIS $y = 0.01$	0.954	0.742
SIS $y = 0.03$	0.974	0.838
SIS $y = 0.1$	0.832	0.887
SIS $y = 0.3$	0.800	0.888

Table 6.4: Comparison for Fitting Factor results between Cao et. al [4] and bilby for point mass and SIS lensing profiles using a lens mass of $1000M_{\odot}$

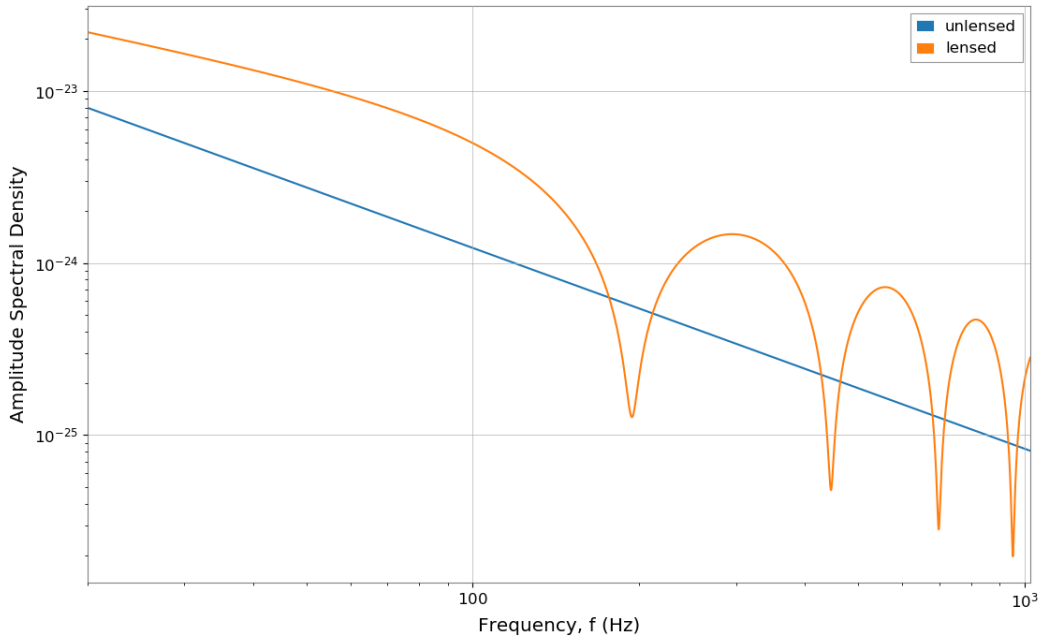

Figure 6.11: Comparison of Amplitude Spectral Density between the lensed and unlensed signals, these were the components for calculating the fitting factor. To mirror Cao et al [4], these signals were generated using two black holes of $10M_{\odot}$ and set at appropriate distance to make the SNR = 10

Chapter 7

Results and Analysis

Taking from Chapter 6, that the code is able to accurately produce and perform analysis of lensed and unlensed gravitational wave signals in cases where y is 0.1 or greater, the code may now be used to investigate signals to determine the extent to which lensed and unlensed signals can be confused for one another, or conversely how well one may be able to determine if a signal possesses a particular lensing profile. The following presents the results of that investigation using simulated data as they would be detected by a combination of LIGO, VIRGO, and KAGRA.

7.1 Parameter Estimation Using Same Model Type

The first thing to show will be the main way that bilby presents its results. These are corner plots. The simplest case of analysis - and a final demonstration of the efficacy of the bilby analysis is to show that given a particular set of parameters for each of the models, bilby analysis is able to return paramters close to the 'true' case when asked to return parameters for that model. This is presented for each of the models in the figures below.

In each case, the base of the model was a binary black hole inspiral with $m_1 = 34$, $m_2 = 29$, at a luminosity distance of $D_L = 410$ Mpc. The binary black hole was placed at $(1.375, 1.2108)$ for the RA - Dec co-ordinate system. Each of the other parameters that may be defined for the system was set to zero. This generated the initial unlensed signal which could then be easily lensed. The lensing mass for each of the profiles was chosen to be $1000M_\odot$ and the impact parameter chosen to be 0.1 for each case apart from the Soliton were it was chosen to be 0.044.

Each of these demonstrates that when a model is given and the correct model is analysed for, that bilby is able to return the correct parameters for the model, with a singular exception - that being the value of y for the NFW case. One reason for this may be that, as can be seen in Figure 6.6, the waveforms at different y values are much more similar than for those of the other model types which would mean that in noisy data,

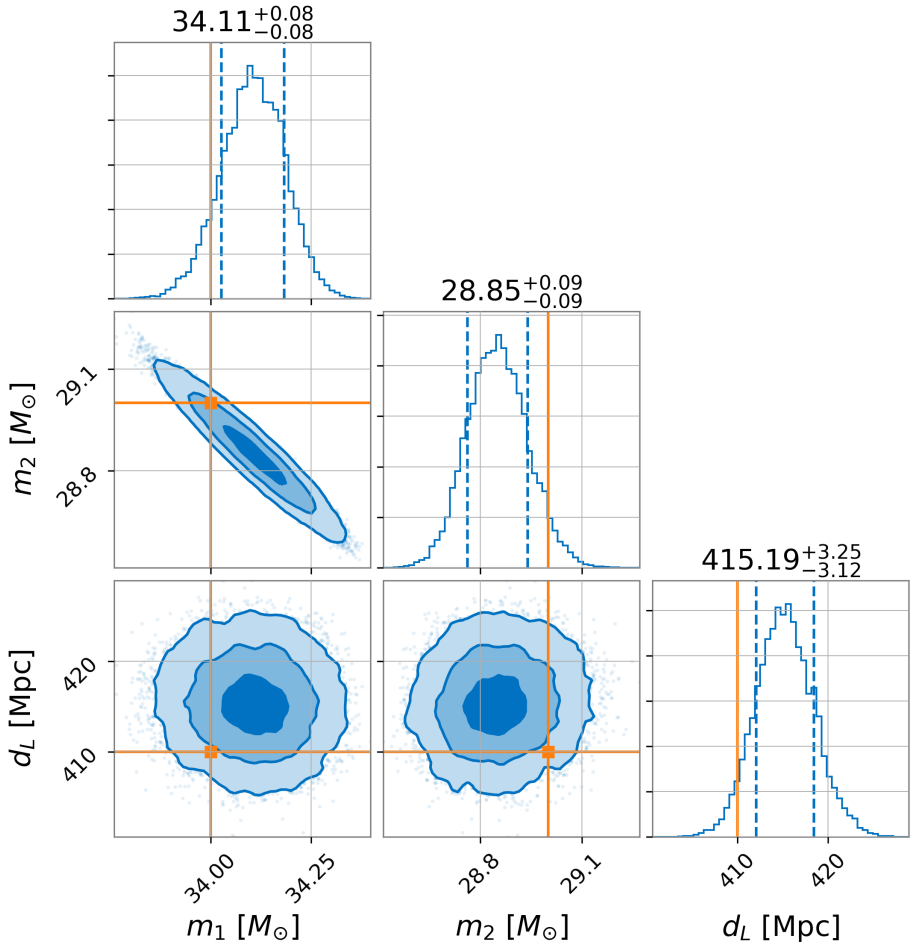

Figure 7.1: Results of bilby nested sampling for parameter estimation of an unlensed gravitational wave signal when treating it as such. The 'true' values are superimposed upon the contours at the intersection of the orange lines with the estimated parameter values given above the relevant column. It can thus be seen that bilby successfully recovers values similar to the 'true' values.

constraining y for the NFW case would be more difficult. However, the remainder of the parameters are able to be well constrained, which lends confidence in using the code in future investigations.

7.2 Parameter Estimation Using Different Model Type

Having established that the bilby analysis returns values close to the 'true' case when investigating a signal using the correct model, it is next worth showing what happens in the case when the wrong model is used - for instance analysing a lensed signal as an unlensed signal, or using the wrong lensing profile for a lensed signal. This analysis wil serve to demonstrate the value in being able to differentiate between the models, by showing how the parameters that are estimated are affected by the wrong model being used. Figures

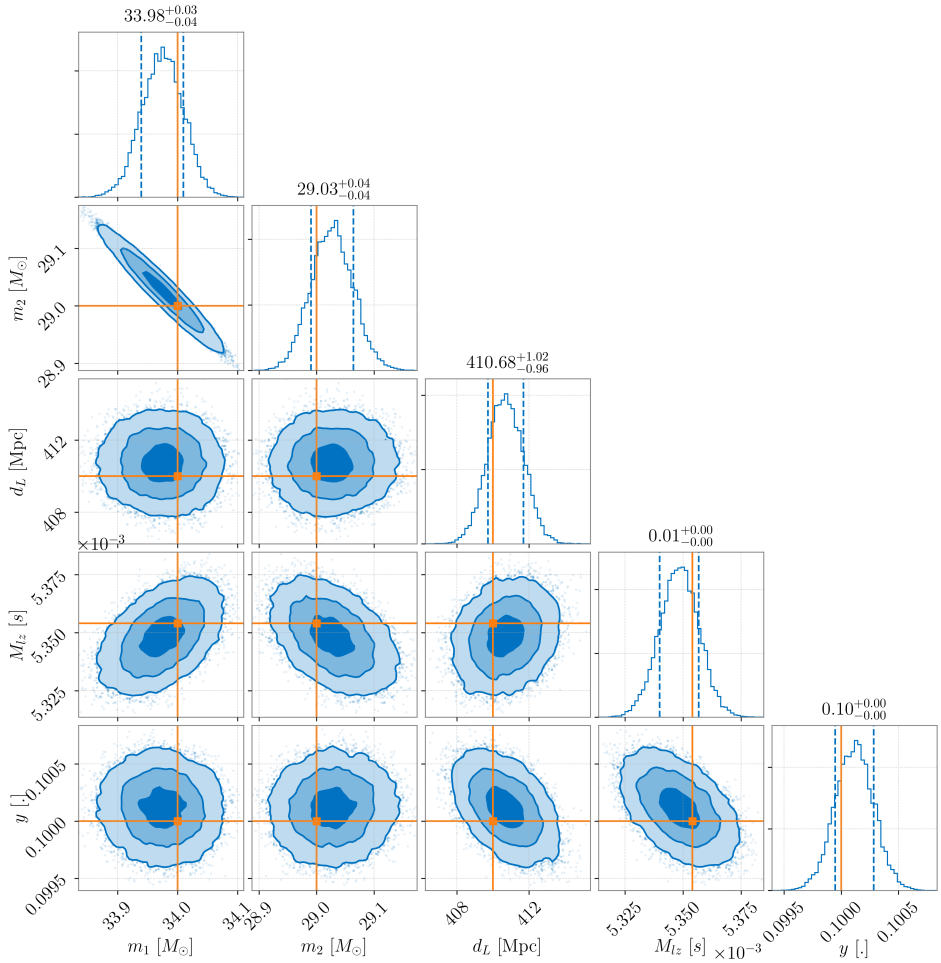

Figure 7.2: Results of bilby nested sampling for parameter estimation of a gravitational wave signal lensed using the Point Mass lensing profile and treating it as such. As can be seen, bilby is once again able to recover estimated values that are close to the 'true' values for both the signal and lens properties examined.

7.6 through 7.9 show the results of treating a lensed signal as unlensed, including showing that the parameters that are unused in the unlensed case - i.e. redshifted lens mass and impact parameter - are not constrained at all in that analysis.

The contours in this case are visually much less well structured - perhaps immediately suggesting that these results are less sensible than the case where the correct model has been used. However, examining the values put forward based upon these contours, these do not necessarily immediately discount the results as invalid. Each of the model yields similar, though different enough to be erroneous - the scale of difference being in the range of 10%, results for the pair of masses. The more dramatic difference is to the luminosity distance where treating the lensed models as unlensed yields a difference between a half and a quarter of the original value depending upon the model that has been used, which is in approximate alignment with the inverse of the value of the amplification factor. This indicates that should a lensed signal be misinterpreted as an unlensed signal, the effect of

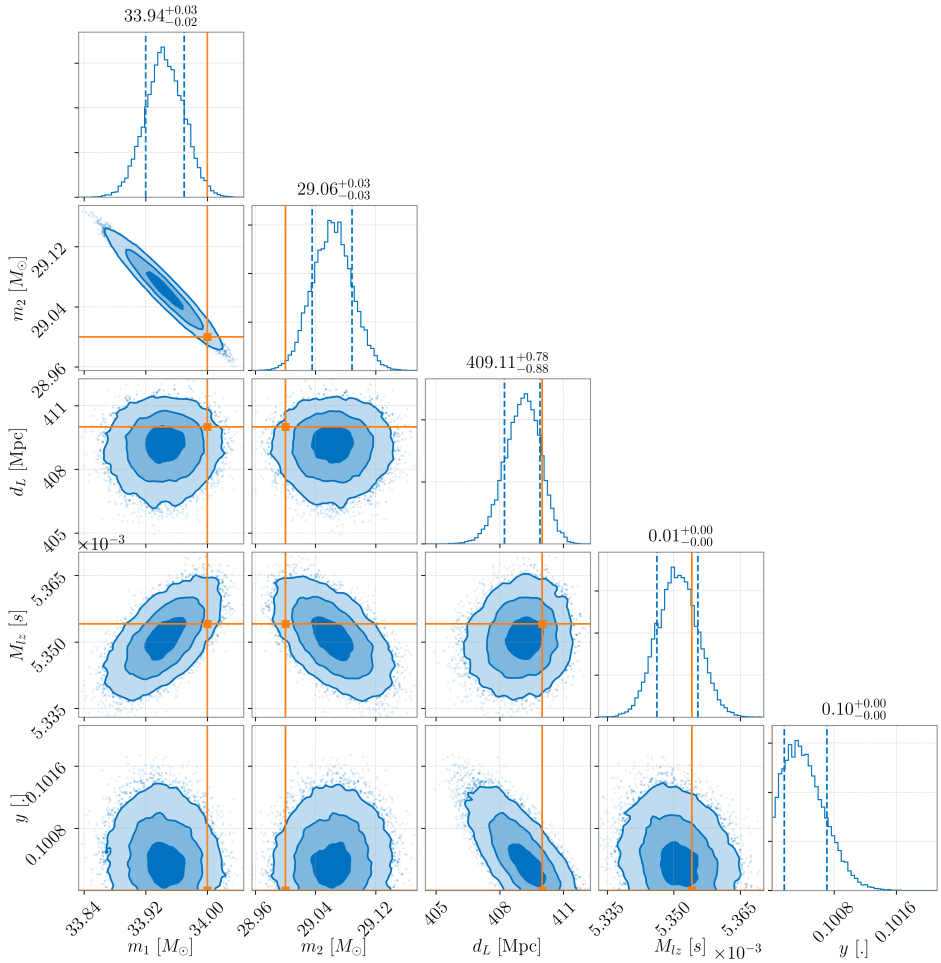

Figure 7.3: Results of bilby nested sampling for parameter estimation of gravitational wave signal lensed using the SIS lensing profile and treating it as such. As can be seen, bilby is once again able to recover estimated values that are close to the 'true' values for both the signal and lens properties examined.

the lensing is to bias the distance estimate to be closer by a factor approximately equal to that of the amplification factor and as such will mean the distance estimate is wildly inaccurate. As expected, the parameters that are not included in the unlensed case - the redshifted lens mass and the impact paramters - have no sensible contouring.

As an initial visual guide before investigating more quantitatively examining the evidences for the model - the corner plots demonstrate the importance of being able to differentiate between the lensed and unlensed cases. If they cannot be differentiated between, then the luminosity distance values for arbitrary signals start to become suspect given the wide division between the true value and the value returned when a lensed signal is treated as unlensed. Given that one of the major benefits of gravitational wave observation is that each merger can yield distances, this would be a significant source of frustration should these differences not be able to be confirmed. However, the lack of structure in the contours when the lensed signals are treated as unlensed does yield an

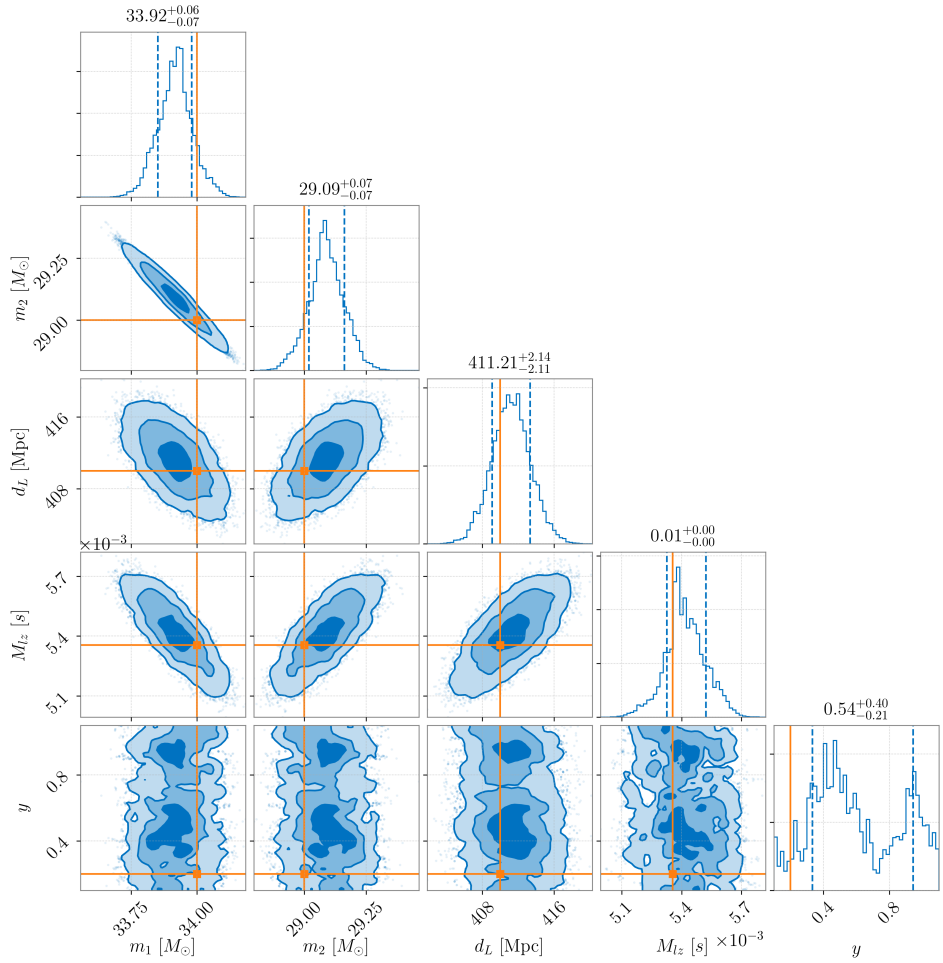

Figure 7.4: Results of bilby nested sampling for parameter estimation of a gravitational wave signal lensed using the NFW lensing profile and treating it as such. As can be seen, for the properties of the gravitational wave signal, bilby once again recovers values similar to the 'true' value. For the lens properties examined, however, the picture is more mixed. The redshifted lens mass is recovered and well constrained, however, for the impact parameter, y, the contours are much more chaotic and are constraining it much less well. As well, note that the value estimated of 0.54 is signifcantly different to the 0.1 'true' value.

early sense that the evidences should be sufficiently lower than those for the case where the signals are treated with the correct model.

Being able to differentiate between only lensed and unlensed is not the only goal being investigated. The follow-up question is whether different lensing models can be differentiated between as well, so it is worth showing the resulting corner plot in the case where a lensed signal is treated as though it were another lens model to investigate how this affects the parameters returned as being the maximum likelihood by the bilby analysis. This is shown in Figure 7.10 below.

In this case, the contours look much more structured again. Once again, the masses returned are similar to the original, though once again, different enough to be erroneous

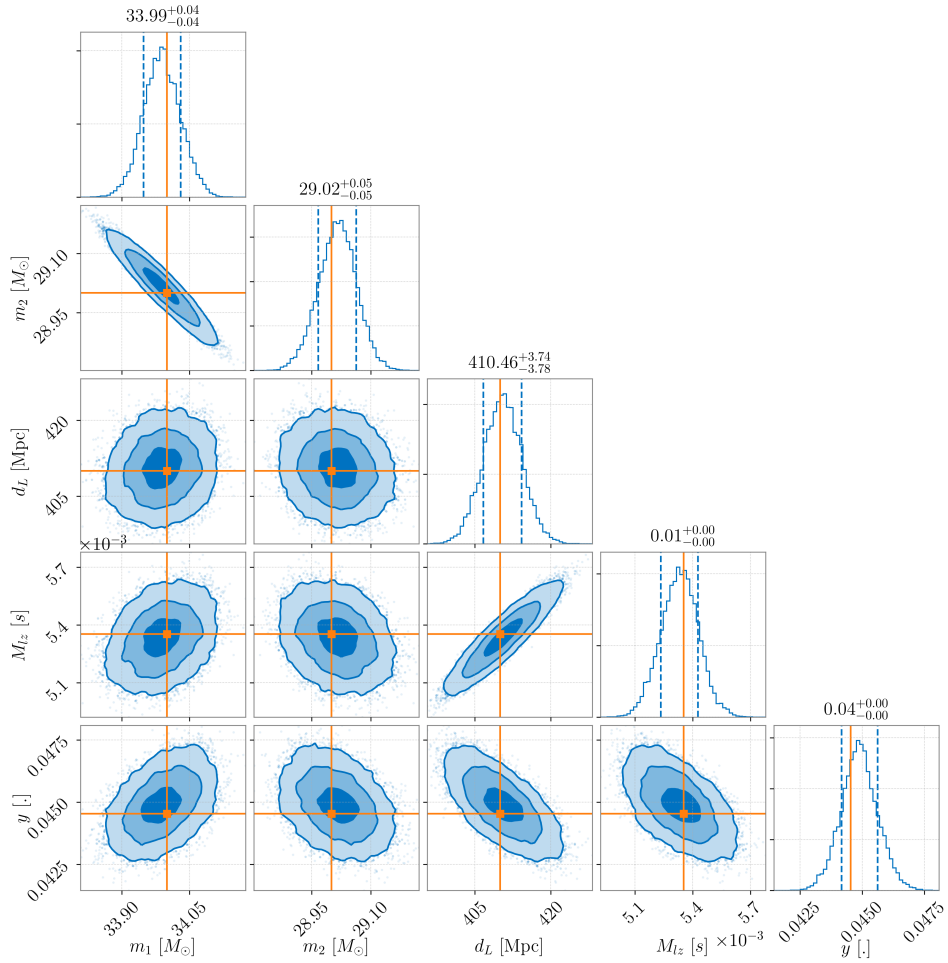

Figure 7.5: Results of bilby nested sampling for parameter estimation of a gravitational wave signal lensed using the Soliton lensing profile and treating it as such. As can be seen, for both the properties of the gravitaitional wave signal and the lens, bilby is able to recover values similar to the 'true' values.

- only m_1's true value is inside any of the contours at all. The luminosity distance in this case, is significantly higher - indicating once again that in order to get accurate distance measurements to a gravitational wave source, the correct modelling needs to be applied. The value for the impact parameter is also significantly higher than the true value. This would give an indication that the task of separating the lensed models is likely more difficult than that for the lensed vs unlensed, and that it would appear that a lensed signal from one profile can appear similar to a lensed signal using another profile with different parameters.

7.3 Analysis for Simulated Data

Having established previously the necessity of being able to determine the correct modelling type to gain insight into the nature of dark matter, what has been shown in the

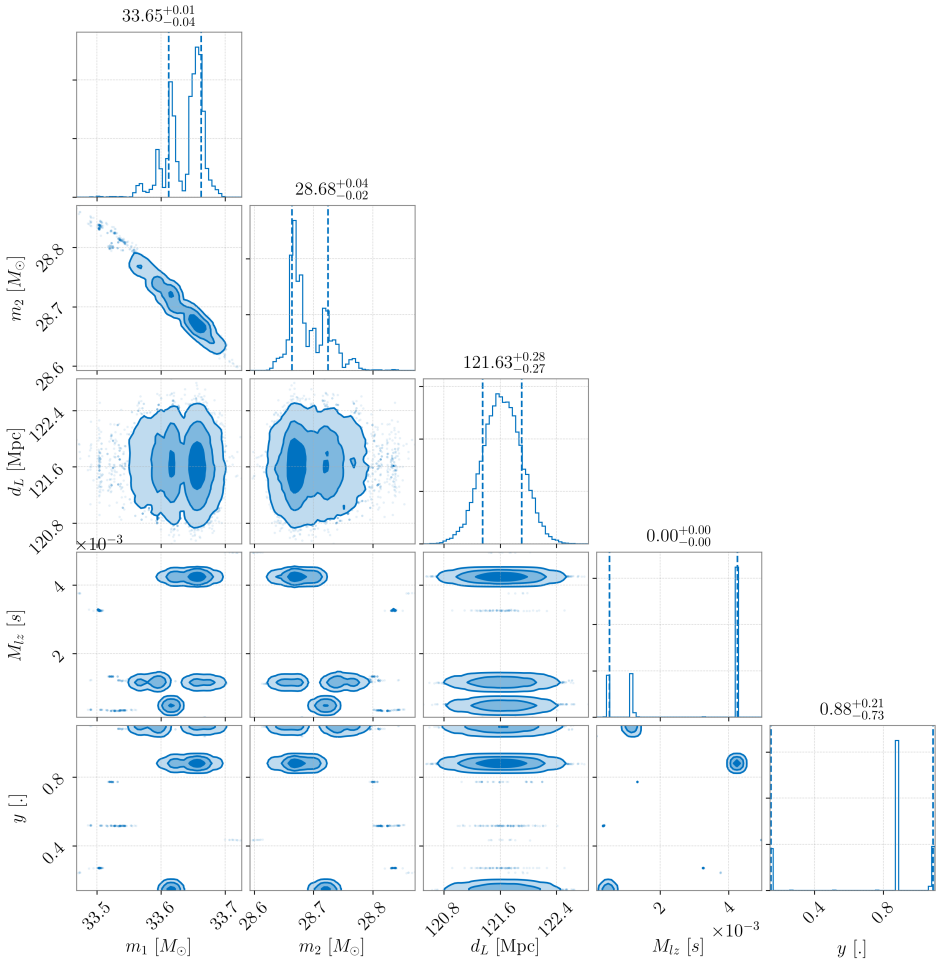

Figure 7.6: Results of bilby nested sampling for parameter estimation for a gravitational wave signal lesned using the Point Mass lensing profile but treating it as an unlensed event to simulate what would happen if lensing is ignored. In this case, the contours are less well behaved which would indicate the parameters are less well behaved. The bilby estimated parameters are also unable to recover values as close to the 'true' values, particularly in the case of the luminosity distance which has been lowered by a factor of 4 in line with the maximal values of the magntiude of the amplification factor.

previous section is that it is also essential for the accuracy of parameter estimation for gravitational wave events as if, for instance, a lensed event is treated as an unlensed event the luminosity distance will be biased towards being closer by a factor equivalent to the magnitude of the amplifaction factor. The analysis here will thus investigate the ability to determine the correct modelling type for realistic simulated data. Given that the method for parameter estimation in the previous sections was nested sampling, the log evidences are calculated during analysis for a given model meaning that the log of the Bayes Factor can be calcuated easily using simple subtraction of those calculated log evidences.

In the following subsections, this analysis will be performed for various parameter sets allowing the investigation not just to see if it is possible to investigate the effects of parameter adjustment - in particular the luminosity distance, and the lens mass on the

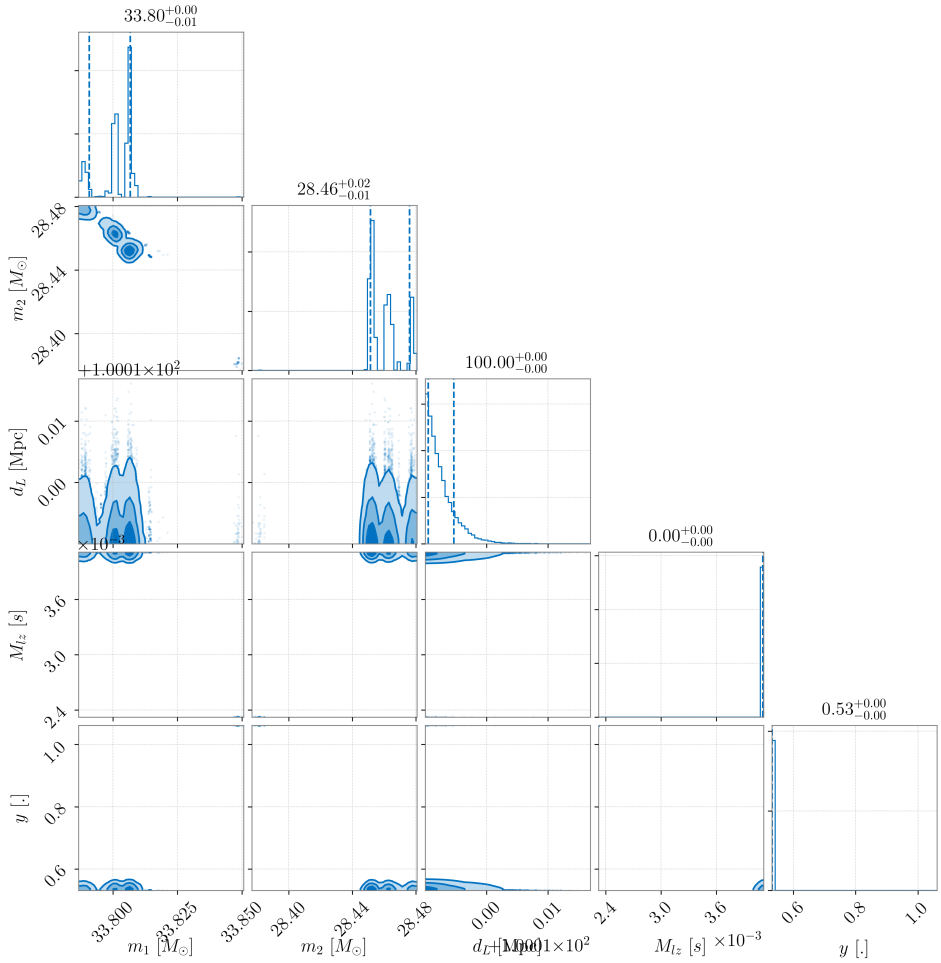

Figure 7.7: Results of bilby nested sampling for parameter estimation of a gravitational wave signal lensed using the SIS lensing profile but modelled as an unlensed signal. Again, bilby is unable to recover values as close to the 'true' values as compared to treating the signal as the correct model type. Once again the luminosity distance is lowered by a factor of 4, again approximately in line with the maximal values of the amplification factor. The parameters unused for the unlensed signal are again unconstrained.

ability to differentiate between them. The investigation by Cao et. al, that was recreated in the previous chapter would initially suggest that a Point Mass lensed signal and an unlensed signal become more similar as the lens mass is decreased for instance.

7.3.1 Case 1: 'Standard' Parameter Set

In the following analysis, a run will be completed for each of the lensed signal types using the following parameter set: $m_1 = 34M_\odot, m_2 = 29M_\odot, D_L = 410Mpc, M_{lens} = 1000M_\odot, y = 0.3$ which will be treated as the standard to which the other cases will be variants of. The Signal-to-Noise Ratio, or SNR, as generated from this parameter set at each of the Hanford, Livingston, and Kagra detectors is given below in Table 7.1. It will be analysed as all of the possible configurations, and the log Bayes factors for each of the

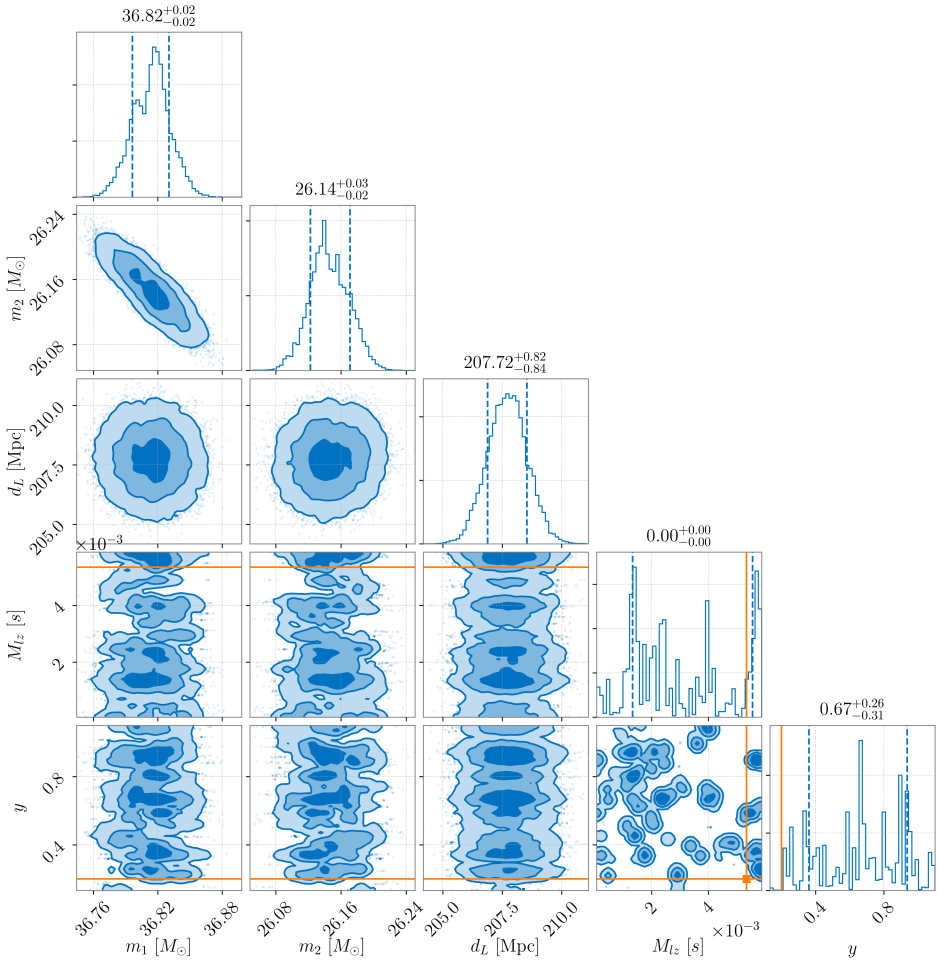

Figure 7.8: Results of bilby nested sampling for parameter estimation of a gravitational wave signal lensed using the NFW lensing profile modelled as an unlensed signal. Again, bilby is unable to recover values as close to the 'true' values as compared to treating the signal as the correct model type. Once again the luminosity distance is lowered by a factor of 4, again approximately in line with the maximal values of the amplification factor. The parameters unused for the unlensed signal are again unconstrained.

other models compared to the true models will be shown. If these are all such that the correct model is favoured over each of the others then this will indicate that it is indeed possible to differentiate between the models. If the log of Bayes factors are however, too close to 0 - this will indicate that differentiating is much more difficult.

Examining Table 7.2, in which the log Bayes Factors have been calculated such that M_1 is the true model - the log Bayes Factors are all significantly positive indicating that in every case, the true model is convincingly preferred over all of the other possible profiles. The pair with the smallest value are the comparison between the Point Mass lens and the SIS lens profiles, indicating that these two are the most likely lens profiles to be mistaken for one another, which is consistent with the prediction from the corner plot discussed in the previous section. In each case, the factors which compare the lens with

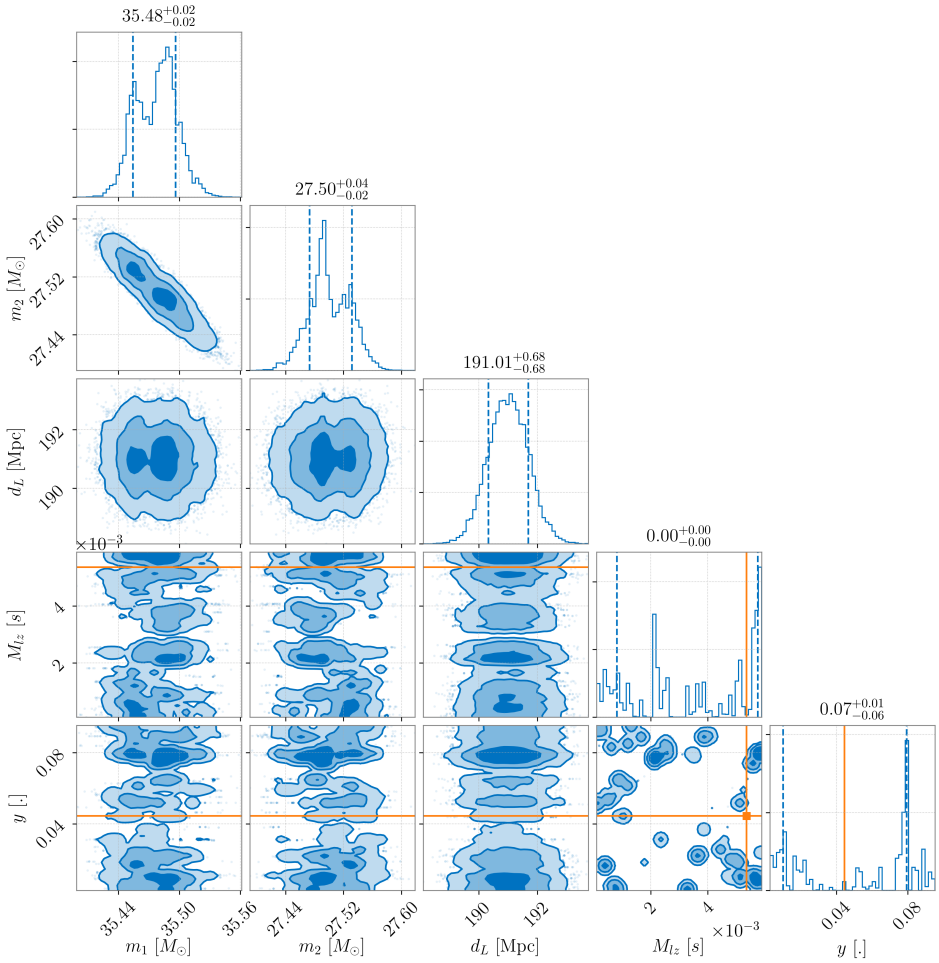

Figure 7.9: Results of bilby nested sampling for parameter estimation of a gravitational wave signal lensed using the Soliton lensing profile. Again, bilby is unable to recover values as close to the 'true' values as compared to treating the signal as the correct model type. Once again the luminosity distance is lowered by a factor of 4, again approximately in line with the maximal values of the amplification factor. The parameters unused for the unlensed signal are again unconstrained.

the unlensed signal type are on the higher end of thpse cacualted - which indicates that lensed and unlensed signals are quite distinct from one another, which again affirms the predictions from the much less structured contours shown in the corner plots resulting from interpreting a lensed signal as an unlensed signal.

7.3.2 Case 2: 10 Solar Mass Lens

The following analysis concerns a completed run using the same inspiral parameters as the previous section, but with a 10 Solar Mass lens as opposed to 1000. Once again, the SNR data is given below in Table 7.3 for the lensed signals - the unlensed signal data being the same as the previous case. These SNR values are significantly lower than the previous case which would give an indication that one should expect less clear data. As

	H1	L1	K1
Unlensed	93.19	85.23	28.74
Point Mass	166.85	152.61	48.28
SIS	160.35	146.60	50.92
NFW	148.44	135.76	46.19
Soliton	205.19	187.67	67.91

Table 7.1: SNR data for each detector for each of the signals. Given that the unlensed signal is used as a basis for the lensed signals the effect of lensing appears to increase the SNR.

	Unlensed	Point Mass	SIS	NFW	Soliton
Point Mass	9024.24	N/A	205.64	9482.87	9197.97
SIS	3543.61	97.96	N/A	3625.73	3542.37
NFW	3249.96	1245.72	1554.43	N/A	2800.86
Soliton	6895.63	1343.95	120.92	1752.70	N/A

Table 7.2: Comparison of log Bayes Factors calculated with M_2 as the top row and M_1 being the left most column. In this case, the value is positive which indicates that the Bayes Factor is greater than one and thus that the probability of M_1 is greater than M_2, indicating that the correct model is correctly identified.

was previously mentioned, in the Cao et al analysis that was partial recreated in Chapter 6, lower mass lenses result in signals that have higher fitting factors when compared with unlensed signals. This means that the signals should be more confusable and thus, at least for those cases, one would expect to see that reflected in the log Bayes Factors - with much more marginal results compared with the 'standard' parameter set above which is being used for comparative purposes.

	H1	L1	K1
Point Mass	103.69	94.83	31.94
SIS	125.41	114.70	39.20
NFW	102.22	93.49	31.47
Soliton	95.19	87.06	29.36

Table 7.3: SNR data for the lensed signals as detected at each of LIGO Hanford and Livingston as well as Kagra for the Low Lens Mass case. Unlensed data is unmodified from the previous case. In comparison with Table 7.1 the values are significantly lower, which would give an indication that the signals may be less distinguishable

As can be seen in Table 7.4, this prediction is very much reflected in what is seen in the results - the log Bayes Factors are far smaller in comparison with the 'standard' case results, and in fact in three cases the values are negative meaning the wrong model is preferred meaning that in this case, the analysis at present is insufficient to differentiate

	Unlensed	Point Mass	SIS	NFW	Soliton
Point Mass	176.76	N/A	49.11	17.92	-0.61
SIS	464.48	14.38	N/A	3.30	18.92
NFW	49.95	9.40	80.57	N/A	6.62
Soliton	3.06	-94.82	329.76	-156.42	N/A

Table 7.4: Comparison of log Bayes Factors calculated with M_2 as the top row and M_1 being the left most column. In this case, some of the entries are negative which indicates that the Bayes Factor is less than one, and that the wrong model is being preferred. This would indicate that for lower lens masses, it is more difficult to distinguish between the models and that the detector setup being simulated is insufficient to prefer the correct model in all cases.

between the different lens types for low mass lenses. Examining the Unlensed comparison, by a small margin, the lens models are still preferred over the unlensed cases indicating however, that examining a signal would slimly yield that the signal was lensed even for small lens masses. This belief would be increased for Point Mass and SIS, however, would be only slight for the Soliton model.

7.3.3 Case 3: Higher Distance

The parameter that is most affected by analysing a signal as the wrong lensing profile is the luminosity distance to the object - which makes investigation worth doing into whether distance affects the ability to differentiate between the different profiles. An initial suspicion may be granted by the fact that the amplitude is inversely proportional to distance - signals from further away are weaker which means that the noise is more impactful and the SNR is lower - this is confirmed by Table 7.5 which contains the lowest SNR values of the three - which one would anticipate would mean that the different profiles may be more easily confusable for one another. To investigate this, the following analysis was done modifying the 'standard' parameter set by doubling the distance to 800 Mpc.

	H1	L1	K1
Unlensed	46.60	42.62	14.37
Point Mass	52.24	47.78	16.09
SIS	63.30	57.89	19.80
NFW	51.22	46.85	15.78
Soliton	47.67	43.60	14.70

Table 7.5: SNR data for the signals as received by the LIGO Hanford and Livingston detectors as well as KAGRA for the higher distance case. As would be expected for the unlensed case, the increased distance to the source has resulted in the lowering of the SNR and this is true of the lensed signals as well. Here, the increase from the unlensed figures is the smallest, giving an indication that this data should be the least distinguishable.

	Unlensed	Point Mass	SIS	NFW	Soliton
Point Mass	-206.71	N/A	9.47	5.38	-2.7
SIS	135.21	4.07	N/A	-0.42	-0.15
NFW	8.54	0.18	13.15	N/A	0.48
Soliton	-90.54	-72.45	10.28	2.19	N/A

Table 7.6: Comparison of log Bayes Factors calculated with M_2 as the top row and M_1 being the left most column. In a similar manner to the previous case, some of the entries are negative indicating that the Bayes Factor is less than one and hence prefers the wrong model, indicating that at higher distances again distinguishing the models is more difficult and the detector setup simulated is insufficent to prefer the correct model in all cases. Directly comparing the number of negative entries with the previous case would also suggest that it is more difficult in the case of higher distance compared with the case of lower lens mass.

As can be seen, this prediction is very much reflected in the results with some of the lowest values for the log Bayes factors and more cases where the values are negative and the wrong model is favoured. These results indicate that the modern gravitational wave detectors whose sensitivity curves were used to help generate the models of received signals are unable to determine the difference between the different lens profiles and between the lensed and unlensed cases at this distance. As mentioned in the previous paragraph - given the lower SNR, this highlights the need to continuously improve the sensitivities of detectors in order to increase the distance at which the profiles are distinguishable.

7.4 Result Summary

In summary, given that Chapter 6 has given justification for the use of the bilby analysis tools for gravitational wave signal analysis. Chapter 7 has covered the actual usage of the code on the problem of gravitaitonal wave lensing - whether an arbitrary signal can first be categorised in terms of lensed or unlensed, and then categorised into individual lensing profiles. It has also given illustraton of the problems that arise if a signal is miscategorised in terms of the estimated parameter values.

What was done first in §7.1 was to demonstrate that in the case of each lensing type being analysed correctly, i.e. the case of the signal being correctly identified, the parameter estimation behaves sensibly. In this case, the corner plots generated (Figures 7.1 - 7.5) show well behaved contours, and each correctly identifies appropriate estimates for the parameters involved, with the true values of the generated data inside the constraints identified. This is with a single exception, for the impact parameter values for the NFW case - in this particular instance, the analysis is not able to well constrain the values. This is reflected in a much more chaotic picture in the contours in Figure 7.4.

What was done next in §7.2 was to illustrate the problems that arise should a signal

be misidentified, for example, should a Point Mass signal be interpreted as an unlensed signal. In this case, what was examined was the case where each of the lensing profiles was misinterpreted as an unlensed signal. Here, what is seen in Figures 7.6 - 7.9 is that the masses of the system are not greatly affected in terms of the estimated values although the peaks in likelihood are much less singular, with the estimates recovered being similar to the true values, however, what is greatly affected is the luminosity distance estimate to the system. This is biased as being closer by a factor approximately equivalent to the maximal values of the amplification factor. What this suggests is that the lensing process affects the amplitiude, where the distance estimate can be made from, much more than it affects the frequency behaviour of the signal, where the masses are interpreted from. Additionally shown in the results is that interpreting a signal as unlensed, a model where no lens is present, causes no constraint in lens parameters as would be expected.

A single additional case was considered in this analysis to illustrate the biases caused by misinterpreting one lensing profile as another. In this case, an SIS lensed signal was treated as though it were a Point Lens signal with the results shown in Figure 7.10. Here, the behaviour of the contours seems much more well behaved again, with sensible appearing results. However, the estimated values again for luminosity distance and impact parameter are both biased to a large degree. In this case, the luminosity distance is biased away from the observer - in opposition to the unlensed vs lensed comparison - by a factor approximately equal to the ratio of the maximal values of the amplification factor. The impact parameter is biased as occurring further from the lensing object.

Having confirmed that the effect of misidentifying a signal type is to greatly distort the parameter estimates, the analysis in §7.3 turns to investigating how easily one can distinguish each of the lens types from each other and from the unlensed case. This is presented for a 'standard' case in Case 1, and then for a much smaller lens in Case 2, and a much more distant system in Case 3. These resulted in the log Bayes factor tables 7.1, 7.2, and 7.3. In the standard case, each of the factors is largely positive - particularly in the case of the left most column which concerns the ability to distinguish each of the lens profiles from unlensed. This would indicate that it is relatively easy to distinguish lensed signals from unlensed ones. Each of the other combinations in the standard case, similarly indicates that the correct model is strongly favoured. This gives an indication that in principle, correct identification is possible.

In the case of lower lens mass, results shown in Table 7.2, - where the effects of lensing would be much weaker, a small number of negative values appear meaning that the wrong case is favoured. In particular, when the Soliton and Point Mass are misidentified as the other, the analysis indicates that the misidentification is favoured over the true model. This is also the case where the Soliton is misidentified as the NFW profile. This indicates that whilst lensed signals may still be distinguished from unlensed signals with relative

ease, given that this column contains some of the more convincingly positive entries, the seperation of individual lens profiles becomes more difficult given the much smaller lens. In the case of higher distance, results shown in Table 7.3, there are more negative values - i.e. more cases in which the signal is misidentified. This case, also includes cases where the signal is misidentified as unlensed - both the Soliton and Point Lensed signals are so misidentified. This case also includes much smaller values for the log Bayes Factor in general, indicating a lesser degree of confidence where the correct profile is identified. This would indicate that at higher distances, even the question of lensed vs unlensed becomes signficantly more difficult. Taken in combination, this would suggest that detectors have a range of lens mass and distance over which they are able to correctly categorise the signals that they receive.

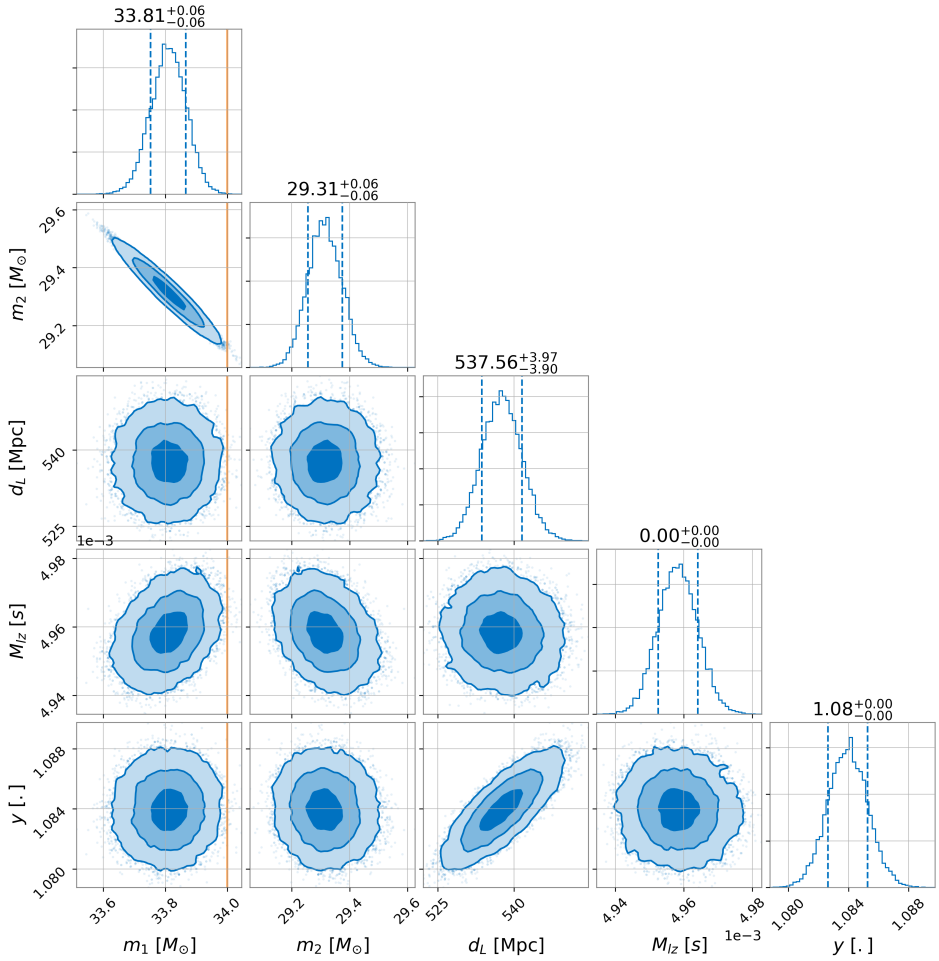

Figure 7.10: Results of bilby nested sampling for parameter estimation for a gravitational wave signal lensed using the SIS lensing profile, but treating it as a Point Mass lensing event. In this case, the parameters of the signal estimated by bilby are closer to the 'true' values when compared to treating a lensed event as an unlensed event. Note, that both the luminosity distance and impact parameter have signifcantly different values, yet more well behaved appearing contours. In this case, the luminosity distance is increased by a factor of $\frac{1}{4}$, which is approximately in line with the ratio of the maximal values of the SIS and Point Lens amplification factors. The impact parameter is increased tenfold compared with the 'true' value of 0.1.

Chapter 8

Conclusion

The observation of gravitational waves has been the result of a century of gravitational theory and experimental design. With the initial problem of *how* to detect them largely solved in the form of modern Laser Interferometric Gravitational Wave Observatories in the vein of LIGO and VIRGO, the focus moves more to the future and what this entirely new method of looking out to the universe can tell us about the unsolved problems of the universe - such as those in Cosmology.

Gravitational Lensing of Gravitational Waves is an important phenomenon in helping to answer one of these unsolved problem. The profile of the lensing that is encoded into the signal is dependent upon the mass density profile of the object that the lensing has been done by and this is key to the methods by which the lensing can help our understanding. Dark Matter has been mysterious precisely by its nature of not interacting electromagnetically but only gravitationally meaning that whilst they cannot be investigated deeply only using the electromagnetic spectrum that we have been restricted to in the past, observing for their impact in the form of Gravitational Lensing of Gravitational Waves allows this investigation to be performed.

The question then turns to how these signals are determined to be either lensed or unlensed, and further how the different lens profiles may be correctly identified. The solution here comes from a tool that has seen a renaissance in the past decades - Bayesian Inference. This presents a means by which the properties of any received signal may be looked at and compared to the expected theoretical properties of any given model and presents a mathematical framework to look at the results. It has been answering this question, that the content of this thesis has largely concerned itself with.

What has been presented here has been an investigation into the use of an implementation of Bayesian style nested sampling using software developed as a combination of Mathematica scripts created by Antonio Hererra Martin to generate the amplification factors for each of the lensing profiles and a tool called 'bilby' implemented in the python programming language to both generate and determine if signals detected gravitational

wave observatories would be able to be categorised definitively into lensed or unlensed - and into the individual lens profile types in the hopes of being able to probe objects in a new manner.

Chapter 6 concerned the question of testing this piece of software. The first test performed was a comparison evaluating the ability of the python implementation of the Point Mass amplification factor generation compared with the original Mathematica, and proving that the more complex methods were unable to be replicated in python successfully requiring the use of the original Mathematica to generate the signals correctly. The secondary test was then to look into bilby to evaluate its ability to generate sensible evidence. The way that this was done was to compare bilby's results to a scenario in which the Bayesian evidence could be analytically evaluted. This toy problem evaluation yielded a good fit between the analytical values and those returned by bilby justifying its use for the task. The final comparison that was performed was a recreation of the analysis performed by Cao et. al [4] examining a fitting factor which allowed for comparison between the generated lensed signals and unlensed signals. This provides a numerical baseline for how close these signals should be. Comparing the recreated results with the original yielded similar behaviour in the sense that larger impact parameter values and lower mass result in more similar signals, the actual numerical values were different. Where the value of the impact parameter was above 0.1 - the values were within 10% of each other, which justified the use of the software in these scenarios.

Chapter 7, having taken it that Chapter 6 was sufficient to justify its usage, performed the analysis of the signals. First generated were the results of performing parameter estimation on each signal type when it was treated as that signal type, this demonstrated the ability of the software to pick out the correct parameters when the correct model was used. What was then done was to generate the results of parameter estimation treating each signal type as the wrong type - showing that the parameters returned were different, with specifically the luminosity distance being extremely different. This showed the necessity of being able to tell the signals apart as in addition to the benefits in terms of new information gained, if the correct model is not used then the parameters estimated will be erroneous. Finally, the Bayes Factors were evaluated for comparing each signal type to each other signal type for three different cases - a 'standard' case, a low mass lens object case, and a high distance case.

The main results yielded were that for the standard case, the Bayes Factors were sufficiently in favour of the correct model in all cases to definitively point towards the correct model - but that this was no longer true in the other two cases - with the high distance case proving particularly narrow with the lowest values of Bayes Factors and the most scenarios in which the wrong model was identified. This highlights one major point - that the framework can correctly separate the lens profiles in theory, but that in practice

this can be done only for certain cases dependent upon the sensitivity of the detector which immediately yields ideas for future work: examining each detector for its sensitivity to lensing and what kinds of lenses it will be able to detect and at what distances it will be able to detect them, and investigating what improvements can be made on a software level to improve efficiency, and to improve the number of profiles that can be translated as this will allow for greater portability and make the software more integrated.

The work presented is not fully complete - concerning itself with only establishing a baseline case and then investigating the effects of two changes those being the luminosity distance to the source of the gravitational wave signal and the mass of the object lensing the signal. In the timescale allotted there was insufficient time to test everything that could be of import - examples of some of these are discussed below. The two that were selected were so chosen primarily due to the assumption that they would be some of the most important effects - and indeed the results do indicate that they do give a significant contribution.

Other parameters of the source or the lens would benefit from additional study. For the lens, the effect of the impact parameter, y, on the Bayes factors from comparing the different lens models has not yet been investigated. Considering the effect of lowering the impact parameter on the amplification factor, one would surmise that the likely effect will be to hinder the ability to distinguish between the different lens models. It is also important to consider effects for the gravitational wave sources, as was shown by [98] the effect of lensing can induce errors into for example estimates of the spin of the system in cases where a strongly lensed high SNR signal is also affected by microlensing, also noted is that this results from a single microlensing event where in actual signals there may be multiple such microlensing sources between the source and detector. [99] also notes that third generation detectors may also be able to detect wave optics effects due to intermediate mass black holes which may also affect the ability to distinguish between differing lens models.

What this work does provide, however, is a framework in which each of these possible effects - if they are able to be simulated - may be tested both in isolation and in combination to investigate their effects on the ability to distinguish between models for simulated signals. If successful, the framework could then be used on arbitrary signals to perform this analysis.

Bibliography

[1] Martin A H 2018

[2] Ashton G, Hübner M, Lasky P D, Talbot C, Ackley K, Biscoveanu S, Chu Q, Divakarla A, Easter P J, Goncharov B, Hernandez Vivanco F, Harms J, Lower M E, Meadors G D, Melchor D, Payne E, Pitkin M D, Powell J, Sarin N, Smith R J E and Thrane E 2019 ApJS **241** 27 (*Preprint* 1811.02042)

[3] LIGO Scientific Collaboration 2018 LIGO Algorithm Library - LALSuite free software (GPL)

[4] Cao Z, Li L F and Wang Y 2014 *Phys. Rev. D* **90**(6) 062003 URL https://link.aps.org/doi/10.1103/PhysRevD.90.062003

[5] Barlow R J 2019 *arXiv e-prints* arXiv:1905.12362 (*Preprint* 1905.12362)

[6] Caves C M, Fuchs C A and Schack R 2002 Phys. Rev. A **65** 022305 (*Preprint* quant-ph/0106133)

[7] Abbott B P, Abbott R, Abbott T D *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. Lett.* **116**(6) 061102 URL https://link.aps.org/doi/10.1103/PhysRevLett.116.061102

[8] Rabinowitz M 2007 *arXiv e-prints* physics/0702155 (*Preprint* physics/0702155)

[9] Apian P 1539 *Cosmographia* (A. Berckmano) URL https://books.google.co.uk/books?id=KPMynQEACAAJ

[10] Grant E 1987 *Isis* **78** 153–173 ISSN 00211753, 15456994 URL http://www.jstor.org/stable/231520

[11] Newton I, Koyre A and Cohen I B 1972 *Philosophiae naturalis principia mathematica* 3rd ed (Harvard University Press [Cambridge, Mass.]) ISBN 0674664752

[12] Kant I 1755 *Allgemeine Naturgeschichte und Theorie des Himmels, oder, Versuch von der Verfassung und dem mechanischen Ursprunge des ganzen Weltgebaudes nach Newtonischen Grundsatzen abgehandelt [microform]* (Johann Friederich Petersen Konigsberg ; Leipzig)

[13] Oort J H 1977 ARA&A **15** 295–362

[14] Bayes T 1763 *Phil. Trans. of the Royal Soc. of London* **53** 370–418

[15] Michelson A A and Morley E W 1887 *American Journal of Science* **Series 3 Vol. 34** 333–345 (*Preprint* `http://www.ajsonline.org/content/s3-34/203/333.full.pdf+html`) URL `http://www.ajsonline.org/content/s3-34/203/333.short`

[16] Einstein A 1905 *Annalen der Physik* **322** 891–921 (*Preprint* `https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19053221004`) URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19053221004`

[17] Einstein A 1916 *Annalen der Physik* **354** 769–822 (*Preprint* `https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19163540702`) URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19163540702`

[18] Tong D 2009 *arXiv e-prints* arXiv:0908.0333 (*Preprint* `0908.0333`)

[19] Misner C W, Thorne K S and Wheeler J A 1973 *Gravitation / Charles W. Misner, Kip S. Thorne, John Archibald Wheeler* (W. H. Freeman San Francisco) ISBN 0716703343 0716703440

[20] Anton H 2010 *Elementary Linear Algebra* (John Wiley & Sons) ISBN 9780470458211 URL `https://books.google.co.uk/books?id=YmcQJoFyZ5gC`

[21] Ni W T 2017 *Genesis of general relativity — A concise exposition* vol 1 pp 85–108

[22] Drory A 2015 *Studies in the History and Philosophy of Modern Physics* **51** 57–67 (*Preprint* `1412.4018`)

[23] Rosser W 1992 *Introductory Special Relativity* (Taylor & Francis) ISBN 9780850668384 URL `https://books.google.co.uk/books?id=zpjBEBbIjAIC`

[24] Kurusingal J 2018 *European Journal of Physics* **39** 055804

[25] Forshaw J and Smith G 2009 *Dynamics and Relativity* Manchester Physics Series (Wiley) ISBN 9780470721520 URL `https://books.google.co.uk/books?id=NW4VUsmRXrkC`

[26] Hughes D W 2001 *Journal of Astronomical History and Heritage* **4** 15–28

[27] D'Abramo G 2017 *arXiv e-prints* arXiv:1711.03833 (*Preprint* 1711.03833)

[28] Fleury P 2019 *arXiv e-prints* arXiv:1902.07287 (*Preprint* 1902.07287)

[29] Taylor B N and Thompson A 2008 *NIST Special Publication 330*

[30] Schutz B 2009 *A First Course in General Relativity*

[31] Will C M 2014 *Living Reviews in Relativity* **17** 4 ISSN 1433-8351 URL https://doi.org/10.12942/lrr-2014-4

[32] Carroll S M 2004 *Spacetime and geometry: An introduction to general relativity* ISBN 0805387323, 9780805387322 URL http://www.slac.stanford.edu/spires/find/books/www?cl=QC6:C37:2004

[33] Gray N 2019 *A Student's Guide to General Relativity* Student's Guides (Cambridge University Press) ISBN 9781107183469 URL https://books.google.co.uk/books?id=3HWIDwAAQBAJ

[34] Rakotomanana L 2018 *Covariance and Gauge Invariance in Continuum Physics: Application to Mechanics, Gravitation, and Electromagnetism* Progress in Mathematical Physics (Springer International Publishing) ISBN 9783319917825 URL https://books.google.co.uk/books?id=zQtjDwAAQBAJ

[35] Schwarzschild K 1999 *arXiv e-prints* physics/9905030 (*Preprint* physics/9905030)

[36] Teukolsky S A 2015 *Classical and Quantum Gravity* **32** 124006 (*Preprint* 1410.2130)

[37] Ellis G F R and van Elst H 1999 Cosmological Models (Cargèse lectures 1998) *NATO Advanced Science Institutes (ASI) Series C* (*NATO Advanced Science Institutes (ASI) Series C* vol 541) ed Lachièze-Rey M pp 1–116 (*Preprint* gr-qc/9812046)

[38] Hobson M, Efstathiou G, P E and Lasenby A 2006 *General Relativity: An Introduction for Physicists* (Cambridge University Press) ISBN 9780521829519 URL https://books.google.co.uk/books?id=5dryXCWR7EIC

[39] Walters S 2016 *arXiv e-prints* arXiv:1608.05752 (*Preprint* 1608.05752)

[40] Loveridge L C 2004 (*Preprint* gr-qc/0401099)

[41] Carroll S M 1997 *arXiv e-prints* gr-qc/9712019 (*Preprint* gr-qc/9712019)

[42] Eisenhart L 2016 *Riemannian Geometry* Princeton Landmarks in Mathematics and Physics (Princeton University Press) ISBN 9781400884216

[43] Einstein A 1918 *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 154-167.*

[44] Einstein A 1916 *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 688-696.*

[45] Blair D, Howell E, Ju L and Zhao C 2012 *Advanced Gravitational Wave Detectors* (Cambridge University Press) ISBN 9780521874298 URL https://books.google.co.uk/books?id=mvVBkgQZrecC

[46] Sir Fred Hoyle S, Hoyle F, Burbidge G and Narlikar J 2000 *A Different Approach to Cosmology: From a Static Universe Through the Big Bang Towards Reality* (Cambridge University Press) ISBN 9780521662239 URL https://books.google.co.uk/books?id=lxzxg6iHc1MC

[47] Ryden B 2017 *Introduction to Cosmology* (Cambridge University Press) ISBN 9781107154834 URL https://books.google.co.uk/books?id=07WSDQAAQBAJ

[48] Roos M 2015 *Introduction to Cosmology* (Wiley) ISBN 9781118923283 URL https://books.google.co.uk/books?id=1h7WBgAAQBAJ

[49] Arbey A and Coupechoux J F 2019 *arXiv e-prints* arXiv:1907.04367 (*Preprint* 1907.04367)

[50] Pitrou C and Pospelov M 2019 *arXiv e-prints* arXiv:1904.07795 (*Preprint* 1904.07795)

[51] Bromm V 2013 *Reports on Progress in Physics* **76** 112901 (*Preprint* 1305.5178)

[52] Liddle A 2015 *An Introduction to Modern Cosmology* (Wiley) ISBN 9781118690253 URL https://books.google.co.uk/books?id=6n64CAAAQBAJ

[53] Hubble E 1929 *Proceedings of the National Academy of Sciences* **15** 168–173 ISSN 0027-8424 (*Preprint* https://www.pnas.org/content/15/3/168.full.pdf) URL https://www.pnas.org/content/15/3/168

[54] Lahav O 2001 Observational Tests for the Cosmological Principle and World Models *NATO ASIC Proc. 565: Structure Formation in the Universe* vol 565 ed Crittenden R G and Turok N G p 131 (*Preprint* astro-ph/0001061)

[55] Sylos Labini F and Baryshev Y V 2010 J. Cosmology Astropart. Phys. **2010** 021 (*Preprint* 1006.0801)

[56] Fixsen D J 2009 ApJ **707** 916–920 (*Preprint* 0911.1955)

[57] Smoot G F, Bennett C L, Kogut A, Wright E L, Aymon J, Boggess N W, Cheng E S, de Amici G, Gulkis S, Hauser M G, Hinshaw G, Jackson P D, Janssen M, Kaita E, Kelsall T, Keegstra P, Lineweaver C, Loewenstein K, Lubin P, Mather J, Meyer S S, Moseley S H, Murdock T, Rokke L, Silverberg R F, Tenorio L, Weiss R and Wilkinson D T 1992 ApJ **396** L1

[58] Friedmann A 1999 *General Relativity and Gravitation* **31** 1991

[59] Uzan J P and Lehoucq R 2001 *European Journal of Physics* **22** 371–384 (*Preprint* physics/0108066)

[60] Turner M S 1999 Cosmological parameters *COSMO-98* (*American Institute of Physics Conference Series* vol 478) ed Caldwell D O pp 113–128 (*Preprint* astro-ph/9904051)

[61] Hogg D W 1999 *arXiv e-prints* astro-ph/9905116 (*Preprint* astro-ph/9905116)

[62] Rose R D, Parker H M, Lowry R A, Kuhlthau A R and Beams J W 1969 Phys. Rev. Lett. **23** 655–658

[63] Riess A G, Filippenko A V, Challis P, Clocchiatti A, Diercks A, Garnavich P M, Gilliland R L, Hogan C J, Jha S, Kirshner R P, Leibundgut B, Phillips M M, Reiss D, Schmidt B P, Schommer R A, Smith R C, Spyromilio J, Stubbs C, Suntzeff N B and Tonry J 1998 AJ **116** 1009–1038 (*Preprint* astro-ph/9805201)

[64] Sofue Y and Rubin V 2001 ARA&A **39** 137–174 (*Preprint* astro-ph/0010594)

[65] Primack J R and Gross M A K 2001 *Hot dark matter in cosmology* pp 287–308

[66] Schumann M 2019 *arXiv e-prints* arXiv:1903.03026 (*Preprint* 1903.03026)

[67] Carr B 2019 *arXiv e-prints* arXiv:1901.07803 (*Preprint* 1901.07803)

[68] Matos T, Guzmán F S, Ureña-López L A and Núñez D 2002 *Scalar Field Dark Matter* p 165

[69] Matos T and Ureña-López L A 2000 *Classical and Quantum Gravity* **17** L75–L81 (*Preprint* astro-ph/0004332)

[70] Matos T and Arturo Ureña-López L 2001 Phys. Rev. D **63** 063506 (*Preprint* astro-ph/0006024)

[71] Planck Collaboration, Ade P A R, Aghanim N, Armitage-Caplan C, Arnaud M *et al.* 2014 A&A **571** A16 (*Preprint* 1303.5076)

[72] Laplace P 1820 *Théorie analytique des probabilités* (Courcier) URL https://books.google.co.uk/books?id=J7gl92QRdZMC

[73] Ash R 2008 *Basic Probability Theory* Dover Books on Mathematics (Dover Publications) ISBN 9780486466286 URL https://books.google.co.uk/books?id=dr9IAwAAQBAJ

[74] Gregory P 2005 *Bayesian Logical Data Analysis for the Physical Sciences* (New York, NY, USA: Cambridge University Press) ISBN 052184150X

[75] Trotta R 2017 *arXiv e-prints* arXiv:1701.01467 (*Preprint* 1701.01467)

[76] Robert C P 2015 *arXiv e-prints* arXiv:1504.01896 (*Preprint* 1504.01896)

[77] Skilling J 2006 *Bayesian Anal.* **1** 833–859 URL https://doi.org/10.1214/06-BA127

[78] Abbott B P, Abbott R, Abbott T D, Abernathy M R, Acernese F, Ackley K, Adams C, Adams T, Addesso P and et al 2016 *Annalen der Physik* **529** 1600209 ISSN 0003-3804 URL http://dx.doi.org/10.1002/andp.201600209

[79] Chen C M, Nester J M and Ni W T 2017 *Chinese Journal of Physics* **55** 142–169 ISSN 0577-9073 URL http://dx.doi.org/10.1016/j.cjph.2016.10.014

[80] Dyson F J 1969 ApJ **156** 529

[81] Michelson A A and Morley E W 1887 *Sidereal Messenger* **6** 306–310

[82] Martynov D V, Hall E D, Abbott B P *et al.* 2016 Phys. Rev. D **93** 112004 (*Preprint* 1604.00439)

[83] Aasi J, Abbott B P, Abbott R, Abbott T, Abernathy M R, Ackley K, Adams C, Adams T, Addesso P and et al 2015 *Classical and Quantum Gravity* **32** 074001 ISSN 1361-6382 URL http://dx.doi.org/10.1088/0264-9381/32/7/074001

[84] Harry G M, Gretarsson A M, Saulson P R, Kittelberger S E, Penn S D, Startin W J, Rowan S, Fejer M M, Crooks D R M, Cagnoli G and et al 2002 *Classical and Quantum Gravity* **19** 897–917 ISSN 1361-6382 URL http://dx.doi.org/10.1088/0264-9381/19/5/305

[85] Dyson F W, Eddington A S and Davidson C 1923 MmRAS **62** A1

[86] Bicak J 2000 Selected solutions of einstein's field equations: their role in general relativity and astrophysics (*Preprint* gr-qc/0004016)

[87] Coles P 2001 Einstein, eddington and the 1919 eclipse (*Preprint* astro-ph/0102462)

[88] De Paolis F, Giordano M, Ingrosso G, Manni L, Nucita A and Strafella F 2016 *Universe* **2** 6 ISSN 2218-1997 URL http://dx.doi.org/10.3390/universe2010006

[89] Schneider P, Ehlers J and Falco E 2013 *Gravitational Lenses* Astronomy and Astrophysics Library (Springer Berlin Heidelberg) ISBN 9783662037584 URL https://books.google.co.uk/books?id=XJ3zCAAAQBAJ

[90] Mollerach S and Roulet E 2002 *Gravitational Lensing and Microlensing* (World Scientific) ISBN 9789810248529 URL https://books.google.co.uk/books?id=PAErrkpBYG0C

[91] Kuijken K 2003 The basics of lensing (*Preprint* astro-ph/0304438)

[92] Brimioulle F, Seitz S, Lerchster M, Bender R and Snigula J 2013 *Monthly Notices of the Royal Astronomical Society* **432** 1046–1102 ISSN 1365-2966 URL http://dx.doi.org/10.1093/mnras/stt525

[93] Park Y and Ferguson H C 2003 *The Astrophysical Journal* **589** L65–L68 ISSN 1538-4357 URL http://dx.doi.org/10.1086/375851

[94] Herrera-Martín A, Hendry M, Gonzalez-Morales A X and Ureña-López L A 2019 *The Astrophysical Journal* **872** 11 ISSN 1538-4357 URL http://dx.doi.org/10.3847/1538-4357/aafaf0

[95] Takahashi R 2004 *Astronomy Astrophysics* **423** 787–792 ISSN 1432-0746 URL http://dx.doi.org/10.1051/0004-6361:20040212

[96] Takahashi R and Nakamura T 2003 *The Astrophysical Journal* **595** 1039–1051 ISSN 1538-4357 URL http://dx.doi.org/10.1086/377430

[97] Bartelmann M 1996 A&A **313** 697–702 (*Preprint* astro-ph/9602053)

[98] Meena A K and Bagla J S 2019 *Monthly Notices of the Royal Astronomical Society* **492** 1127–1134 ISSN 1365-2966 URL http://dx.doi.org/10.1093/mnras/stz3509

[99] Hannuksela O A, Haris K, Ng K K Y, Kumar S, Mehta A K, Keitel D, Li T G F and Ajith P 2019 *The Astrophysical Journal* **874** L2 ISSN 2041-8213 URL http://dx.doi.org/10.3847/2041-8213/ab0c0f