Veitch, John D. (2007) *Applications of Markov Chain Monte Carlo methods to continuous gravitational wave data analysis.* PhD thesis.

http://theses.gla.ac.uk/35/

Applications of Markov Chain Monte Carlo Methods to Continuous Gravitational Wave Data Analysis

John D. Veitch

Department of Physics and Astronomy

University of Glasgow, University Avenue, Glasgow, G12 8QQ

Presented as a thesis for the degree of Ph. D.

in the University of Glasgow

Declaration

This thesis describes the work undertaken between October 2003 and December 2006 in the field of gravitational wave data analysis in the LIGO Scientific Collaboration (LSC). As such, it depends on the work of the numerous members of the LSC in the development and construction of the interferometers and associated apparatus for data acquisition and analysis. My work was also informed by the existing data analysis algorithms, many of which were developed within the LSC. The work described within was carried out by myself, with the help of my supervisor Graham Woan and my second supervisor Martin Hendry. Some of the work was produced in collaboration with members of external institutions, and of the LSC, and this is noted where applicable in the text.

In the first chapter, I provide an introduction to the nature of gravitational waves, their generation and effects, and some material on their detection. This work is derived from the literature, and previous results from the LSC are included.

The second chapter provides an overview of Bayesian inference, concentrating on the main points which are used throughout the thesis. This is mainly derived from the literature. A description of the algorithm which was developed is given, including techniques used to tune its performance. This work was carried out in collaboration with others.

In chapters three and four, I describe the results of applying the algorithm to first synthetic and then observational data from the LIGO Hanford interferometer. The analyses and development of a means of setting an upper limit were performed by myself.

Chapter five describes the extension of the reversible jump MCMC algorithm, developed by others and implemented by Umstätter *et al.* [1, 2] The work on adding the estimation of sky position, and on the derivation of the approximated Fisher matrix was undertaken by myself.

Disclaimer

In accordance with the LIGO Scientific Collaboration Publication and Presentation Policy, I declare that the results in Chapter 4 are under review and subject

to changes in the future. In particular, the upper limit found for gravitational radiation from the remnant of SN1987a has not been reviewed and does not reflect the scientific opinion of the LSC.

Acknowledgements

When I began my time at the Institute of Gravitational Research at Glasgow I was excited by the possibility of doing research on an area of physics that would soon develop into branch of astronomy, and by the cutting edge work that was being done in improving our understanding of gravity. At first I was wary of the statistical nature of the work, but thanks to my supervisor Graham Woan, the power and beautiful simplicity of Bayesian inference soon became clear to me and I have been fascinated ever since. The detection of gravitational waves is a project that has seemingly involved the latest developments in every branch of physics, from laser optics to solid state physics to numerical relativity. It is therefore not surprising that the analysis of the project's data should involve the latest developments in statistical inference, and I have been lucky to be part of that research.

In my time as a PhD student, I have received support, encouragement and distraction from a large number of people, both at Glasgow University and elsewhere, and to them all I extend my greatest appreciation. I would particularly like to thank my supervisor Graham Woan for his patience and uncanny ability to cut to the bottom of any problem which stumped me. Professor Jim Hough, for first employing me as a summer student back in 2001, when I first became involved in the IGR. To my first officemates Réjean and Matt, thank you for welcoming me into the field and providing a base for and continuous help with my work; and to subsequent 465ers Jen and James for continuing the reputation of room 465, long may it continue! For valuable insight and discussions, Martin Hendry and Luis Theodoro. Since the days of my undergraduate degree Bob Taylor, Kenny Walaron, Richard Codling, Helen McGlone and Neil Thompson have been my long-suffering companions, and I wish them all the best of luck with their post-doctoral careers. I'm indebted to Chris for forming Look Up For Danger, and along with Ellie and Matt making my writing up period a lot more fun. To the rest of the physics gang, Siong, Hazel, Ross, Bryan, Dave, Morag, Peter: cheers!

If it weren't for my non-physicist friends I'd probably have gone mad over the last 4 years, so for keeping me on the right path and being the best group of friends I believe I will ever know, thanks to Paul, Mark, Ewan, Helen, Bentley,

Abstract

A new algorithm for the analysis of gravitational wave data from rapidly rotating neutron stars has been developed. The work is based on the Markov Chain Monte Carlo algorithm and features enhancements specifically targeted to this problem. The algorithm is tested on both synthetic data and hardware injections in the LIGO Hanford interferometer during its third science run ("S3"). By utilising the features of this probabilistic algorithm a search is performed for a rotating neutron star in the remnant of SN1987A within in frequency window of 4 Hz and a spindown window of 2×10^{-10} Hz s^{-1}. A method for setting upper limits is described and used on this data in the absence of a detection setting an upper limit on strain of 7.3×10^{-23}.

A further application of MCMC methods is made in the area of data analysis for the proposed LISA mission. An algorithm is developed to simultaneously estimate the number of sources and their parameters in a noisy data stream using reversible jump MCMC. An extension is made to estimate the position in the sky of a source and this is further improved by the implementation of a fast approximate calculation of the covariance matrix to enhance acceptance rates. This new algorithm is also tested upon synthetic data and the results are presented here.

Conclusions are drawn from the results of this work, and comments are made

on the development of MCMC algorithms within the field of gravitational wave data analysis, with a view to their increasing usage.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1   Background

The development of physics and astronomy has shown that conditions in the universe are often far removed from everyday experience, and by investigating these conditions we may test competing hypotheses for their applicability to the broadest possible range of phenomena. Indeed, much of modern physics is devoted to observing the most energetic events known, in the laboratory or in the night sky, in order to probe the limitations of our understanding and thereby make progress.

The General Theory of Relativity is an example of a theory which deals with such extremes, and explains the behaviour of matter and energy in conditions of great density and pressure, such as those that exist in neutron stars, black holes and in the evolution of the universe itself. It describes gravity in terms of the geometry of spacetime, how this is affected by the presence of matter, and how in turn the dynamics of material bodies are influenced by the geometry. The motivation for this came from the problem of reconciling Newtonian gravity, which

describes gravity as acting instantaneously across space, and Special Relativity, which forbids any physical interaction propagating faster than light, if one is to retain causality. In the 1915 General Theory, Einstein was able to give a geometric description of gravity that did not travel instantaneously, and recognised the implication of the existence of gravitational waves.

In spite of this early recognition, Einstein himself doubted that gravitational waves could ever be observed in an experiment, since the gravitational interaction with matter is so small. Today though, there are several searches for gravitational waves under way, of which I am involved in the Laser interferometric Gravitational Wave Observatory (LIGO) and GEO experiments.

This chapter will present a brief introduction to the theory of gravitational waves in General Relativity. For a much more detailed explanation and discussion of the material presented here, consult references [3] and [4] and the article by Thorne in [5].

1.2 Introduction to Gravitational Waves

The description of spacetime in General Relativity (henceforth "GR") follows from Special Relativity, in that space and time are not distinct entities in themselves, but are combined as a 4-dimensional manifold (t, x, y, z), upon which a metric tensor $g_{\mu\nu}$ is defined. In flat spacetime, as in Special Relativity, this is the Minkowski metric with $g_{\mu\mu} = \{-1, 1, 1, 1\}$, $g_{\mu\nu} = 0$ ($\mu \neq \nu$), in units where $G = c = 1$.

However, in GR the manifold may possess a curvature, described by the Riemann curvature tensor $R_{\alpha\beta\gamma\delta}$, and contracted to form the Ricci tensor $R_{\alpha\beta} \equiv R^{\mu}_{\alpha\mu\beta}$. The effect of this curvature is to produce geodesic deviation, such that an inertial

observer will experience deflection of their spacetime path. In this way, what Newton viewed as an instantaneous force, Einstein described as an effect of geometry, where gravitation is determined purely by the local curvature of spacetime.

The source of the gravitational field in GR is the stress-energy tensor $T_{\mu\nu}$, which is a co-ordinate independent object encoding the density of energy and momentum at each point on the space-time manifold. To an observer in a particular co-ordinate system, the components of this tensor are interpreted as the flux of energy-momentum in each co-ordinate direction. For example, the T_{00} component contains the mass density, the source of classical Newtonian gravity.

The connection between the stress-energy tensor and the curvature tensor is given by the *Einstein Field Equation*

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu}, \tag{1.1}$$

where $R = g_{\mu\nu} R^{\mu\nu}$ is the Ricci curvature scalar.

1.2.1 Derivation of Gravitational Waves

For the derivation of gravitational waves, it is convenient to take the weak-field approximation of the theory; since a flat spacetime results in no gravitational field, the weak field is represented as a small perturbation to the flat Minkowski metric $\eta_{\mu\nu}$:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{1.2}$$

where $|h_{\mu\nu}| \ll 1$, following the notation and procedure of [3]. It can be shown that under a Lorentz transform Λ^{α}_{β}, $h_{\mu\nu}$ itself behaves like a tensor, although not

under a general transformation. From this is defined the trace $h = h^\mu_\mu$, and the "trace-reversed form"

$$\bar{h}^{\mu\nu} = h^{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}h. \tag{1.3}$$

It is then always possible to choose a gauge such that $G^{\mu\nu} = -\frac{1}{2}\Box\bar{h}^{\mu\nu}$, so the Einstein Equations can then be written

$$\Box\bar{h}^{\mu\nu} = -16\pi T^{\mu\nu}. \tag{1.4}$$

In empty space, this reduces to the familiar wave equation $\Box\bar{h}^{\mu\nu} \equiv \bar{h}^\alpha_{\mu\nu,\alpha} = 0$, with the simplest solution being the plane gravitational wave,

$$\bar{h}^{\mu\nu} = A^{\mu\nu}\exp\left(-ik_\alpha x^\alpha\right). \tag{1.5}$$

It is shown in Schutz [3] that this wave follows a null geodesic, travelling at the speed of light, and in the transverse-traceless gauge that the amplitude tensor is given by the expression

$$A^{TT}_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & A_{xx} & A_{xy} & 0 \\ 0 & A_{xy} & -A_{xx} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{1.6}$$

for a wave travelling in the z direction. Gravitational waves therefore are transverse waves, with two polarisations corresponding to A_{xx} and A_{xy}. To examine the effect of a gravitational wave on matter as it passes, consider two particles separated by a small distance ϵ in the x direction only, and initially at rest relative to each

other. The proper distance δl between them is given by

$$
\begin{aligned}
\delta l &= \int |g_{\mu\nu} dx^\mu dx^\nu|^{\frac{1}{2}} \\
&= \int_0^\epsilon |g_{xx}|^{\frac{1}{2}} dx \\
&\approx \left(1 + \frac{1}{2} h_{xx}^{TT}\right) \epsilon
\end{aligned}
$$

In the presence of a gravitational plane wave travelling in the z direction, with a non-zero component in the A_{xx} polarisation, this separation becomes

$$
\delta l = \epsilon + \frac{\epsilon}{2} A_{xx} \exp\left(-ik_0 t\right) \tag{1.7}
$$

and so the effect of the incoming wave is to change the physical distance between the two masses (as distinct from the co-ordinate distance which does not change [4]). The change in distance is dependent on the amplitude of the wave, and also the direction of separation of the two test particles. This feature is a result of the polarisation states of a gravitational wave, which will now be illustrated.

1.2.2 Polarisation of Gravitational Waves

Like their electromagnetic counterparts, gravitational waves can be expressed as having either linear or circular polarisation bases. For the linear case, it is traditional to depict the action of the gravitational wave as it acts on a ring of test masses placed in the plane transverse to its direction of travel. In figure 1.1, 21 particles are placed in a stationary ring centred on the origin, parametrised by the 4-vector $\{0, \epsilon \cos\theta, \epsilon \sin\theta, 0\}$. The displacement from the origin is then given by
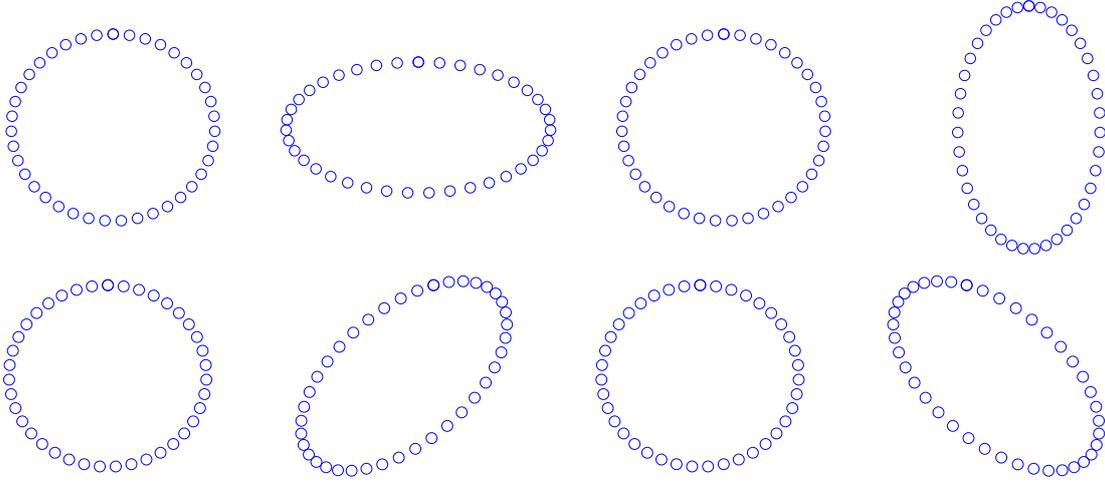
Figure 1.1: The "+" (top) and "×" (bottom) polarisations of a gravitational wave, with increasing time towards the right of the page.

the same technique used above;

$$\delta x \;=\; \epsilon \cos\theta \left(1 + \frac{1}{2} A_{xx} \exp\left(-ik_0 t\right) \right) + \epsilon \sin\theta \left(\frac{1}{2} A_{xy} \exp\left(-ik_0 t\right) \right) \quad (1.8)$$

$$\delta y \;=\; \epsilon \sin\theta \left(1 - \frac{1}{2} A_{xx} \exp\left(-ik_0 t\right) \right) + \epsilon \cos\theta \left(\frac{1}{2} A_{xy} \exp\left(-ik_0 t\right) \right) \quad (1.9)$$

In the upper part of figure 1.1, a wave with only $A_{xx} \neq 0$ is shown, the deformation is in a quadrupolar pattern oriented vertically and horizontally, and is therefore known as the "+" polarisation. The second polarisation state shown in the lower part of 1.1, and has the same form rotated by 45°, known as the "×" polarisation. As the wave passes through, the stress on the ring reverses sign and the contraction on one axis becomes an expansion. This 45° rotation between polarisations, contrasted with the 90° rotation for electromagnetic waves reflects the general fact that the inclination between polarisations of a radiation field of spin S is $90°/S$ [4].

Having shown that gravitational waves are indeed real solutions within the vacuum Einstein equations, it is necessary to enquire about their origins, what systems will produce them, and what information they may carry about their source.

1.2.3 Generation of Gravitational Waves

In this section the generation of gravitational waves will be derived, with the quadrupole mass distribution as the source. The quadrupole moment is the second moment of the mass distribution,

$$I_{jk} = \int T_{00} x_j x_k dV, \tag{1.10}$$

which is not conserved in a general process, and is therefore the lowest order moment which may emit gravitational radiation. Specifically, it can be shown as in [4] that the amplitude of the gravitational wave in the slow-motion approximation (where $v \ll c$ and $h \ll 1$) is related to the quadrupole moment by

$$h_{jk} = \frac{2G}{c^4 r} \ddot{\mathit{I}}_{jk} \tag{1.11}$$

where r is the distance to the source, and I_{jk} is the *reduced quadrupole moment*, obtained by removing the trace from the quadrupole moment:

$$\mathit{I}_{jk} = I_{jk} - \frac{1}{3}\delta_{jk}\text{trace}\,(I_{jk}) = \int T_{00} \left(x_j x_k - \frac{1}{3}\delta_{jk} r^2 \right) dV. \tag{1.12}$$

From the constant of proportionality $G/c^4 \approx 8.26 \times 10^{-45}\,\mathrm{m}^{-1}\mathrm{kg}^{-1}\mathrm{s}^2$ in equation 1.11, it is clear that only very dense objects undergoing acceleration will produce gravitational waves of significant amplitude that might be detected far from the source. For example, consider a simple binary system of two neutron stars, each of mass M, with a circular orbit of radius R and orbital frequency f. The co-ordinate system is defined such that the plane of the orbit is the $x - y$ plane, and at $t = 0$ the stars lie at positions $x = \{+R, -R\}$. If we approximate the stars as point masses, equation 1.12 becomes

$$
\begin{aligned}
\mathcal{I}_{xx} &= 2MR^2\left(\cos^2\left(2\pi ft\right) - \frac{1}{3}\right) \\
\mathcal{I}_{yy} &= 2MR^2\left(\sin^2\left(2\pi ft\right) - \frac{1}{3}\right) \\
\mathcal{I}_{xy} &= MR^2\sin\left(4\pi ft\right) \\
\mathcal{I}_{zz} &= \mathcal{I}_{zx} = \mathcal{I}_{zy} = 0
\end{aligned}
$$

By taking the second time derivatives of these quantities and substituting in equation 1.11, we arrive at expressions for the gravitational wave strain at distance r along the z-axis:

$$
\begin{aligned}
h_{xx} &= -\frac{A_0}{r}\cos 4\pi ft \\
h_{yy} &= \frac{A_0}{r}\cos 4\pi ft \\
h_{xy} &= -\frac{A_0}{r}\sin 4\pi ft
\end{aligned}
$$

where A_0 is the amplitude term

$$A_0 = \frac{32G}{c^4}\pi^2 f^2 M R^2. \tag{1.13}$$

From these equations it is clear that gravitational radiation is emitted at twice the rotation frequency of the binary system, and this is true in general for any quadrupolar gravitational radiation.

From Kepler's 3rd law as written for a circular binary system,

$$4\pi^2 f^2 r^3 = G(m_1 + m_2) \tag{1.14}$$

the orbital frequency and the separation of the stars are related. For an example of a late-stage binary inspiral of two neutron stars, each having mass equal to the sun $(M_\odot = 1.989 \times 10^{30}\,\text{kg})$[6] and a separation of 50 km, the orbital frequency is $232\,\text{Hz}$ (gravitational wave frequency $464\,\text{Hz}$). If the system is situated at a distance of $10\,\text{Mpc}$ along the z-axis, the gravitational wave strain amplitude on Earth will be approximately 2.26×10^{-21}. Although this is a tiny quantity, it represents a significant amount of power emitted by the source. Using the expression for gravitational wave luminosity

$$L_{GW} = \frac{1}{5}\frac{G}{c^5}\left\langle\left(\frac{\partial^3 I}{\partial t^3}\right)^2\right\rangle \tag{1.15}$$

with equation 1.12 gives a total power radiated of 4.183×10^{46} W, and a power flux at Earth of approximately $0.34\,\text{Wm}^{-2}$. The characteristics of this source would make it a candidate for detection by the current generation of laser interferometer gravitational wave detectors, lying in the sensitive frequency range and with suffi-

cient amplitude and duration to have a strong signal to noise ratio. In reality, the orbits of a binary system such as this will rapidly decay due to this emission of energy, producing a strong burst of gravitational radiation upon the final merging of the two stars; and for a time afterward the resulting body would lose energy in this manner too. These are examples of some of the ways in which an astronomical system may produce gravitational waves during the most violent stages of its lifetime.

1.3   Sources Of Gravitational Waves

I will now give a short list of astronomical systems that represent likely sources of gravitational waves. The reason that gravitational waves generated on Earth are not likely to be detectable is apparent in equation 1.13, where the factor G/c^4 must be countered by a very large mass M, which is not possible in the laboratory. There are three main categories into which we can divide the sources of gravitational waves: transient sources, continuous wave sources and stochastic sources.

1.3.1   Transient Sources

Transient sources are those which appear once, emitting a burst of gravitational waves, before fading rapidly. Because they are not repeated, they must produce a high signal to noise ratio in their short lifetime in order to allow a confident detection. Additional evidence, such as their observation in multiple detectors or synchronicity with an observed electromagnetic event may also bolster confidence in a detection. The cause of a gravitational wave burst must be a very powerful

event in order to produce enough radiation to register with detectors on Earth, some possible events of sufficient power are described below.

Supernovae

When the core of a large star undergoes gravitational collapse at the end of its lifetime, a large amount of energy is liberated in a short interval. The detailed process of core collapse is poorly understood due to the difficulty in modelling the full nuclear, stellar and gravitational physics involved in the creation of a neutron star or black hole, which are the end products of this process. Electromagnetic radiation produced in the core is obscured from view by the outer layers of the star, and even the neutrinos generated are scattered by the intervening matter as they heat it. For this reason, by observing gravitational waves from the collapse we stand to gain a great deal of information of the internals of a supernova that cannot be obtained any other way at present. Also due to the difficulty in modelling, the waveforms generated are unknown, and it is not clear whether significant amounts of gravitational radiation will be produced as this depends on deviations from spherical symmetry as the mass distribution changes. However, due to the conservation of angular momentum, the core will be very rapidly rotating, and any instability could potentially lead to such an asymmetry.

If the configuration of the final neutron star or black hole is not spherically symmetric, the compact object will ring down by emitting gravitational waves and becoming smoother.

Binary Inspirals

Compact stars, i.e. white dwarfs, neutron stars and black holes, are expected to exist as binary systems in large numbers in the Milky Way, and these will produce continuous gravitational radiation which causes their orbits to slowly decay in the manner described in 1.2.3 and 1.3.2. As the orbital frequency gets higher and the separation of the stars shrinks, the power radiated grows accordingly and the strength of the gravitational wave grows.

Once such a binary has reached the final few minutes of its lifetime, the behaviour of the orbit becomes highly relativistic. In this regime the post-Newtonian approximation is used to make calculations of parametrised waveforms, allowing the use of matched filtering to perform a search. However, the final plunge stage is not amenable to this treatment, and numerical relativity techniques involving complex simulations must be used to model the large burst of gravitational radiation which is emitted.

During the late-stage inspiral regime, the waveform produced by the pair of stars gradually sweeps upward in frequency producing a "chirp" signal that has been searched for in LIGO data from the S3 and S4 runs, and in the absence of detection a 90% upper limit has been placed on the rate of binary neutron star mergers at < 1.2 per year per L_{10}, on stellar mass binary black holes of < 0.5 per year per L_{10} and of primoridal black holes of 4.9 per year per L_{10}, where L_{10} represents 10^{10} times the luminosity of the sun in blue light [7].

1.3.2 Continuous Wave Sources

Unlike transient sources, continuous wave sources persist over a long period of time while maintaining the same frequency, or with frequency evolving slowly. They are often generated by the rotational motion of a quadrupolar system which is relatively isolated, leading to an approximately stationary frequency which is twice the rotational frequency of the system, as in 1.2.1. The persistence of their signals means that they may be observed over a long period of time, allowing a large amount of data to be collected to improve the signal to noise ratio and estimate the parameters of the source. The statistical methods of this type of analysis shall be introduced in chapter 2, which are applied in this text to two types of continuous sources, neutron stars and binary star systems.

Neutron Stars

A neutron star, being the extremely dense collapsed core of a former star, is a good candidate for the production of gravitational waves. As the angular momentum of the original star is conserved, a neutron star will rotate rapidly unless damped, and many rotating neutron stars have been observed as pulsars with rotation frequencies as high as 716 Hz [8]. This rotation may produce continuous gravitational waves when combined with an asymmetric distribution of mass, or free precession in the star. Although not a continuous source, it is also possible that a starquake, or cracking of the crust of a neutron star, may produce a burst of gravitational waves by excitation of the quasi-normal modes of oscillation [9].

In my work I have concentrated on the emission of gravitational waves from an asymmetric distribution of mass under rotation about an axis; I will take a closer

look at how this arises in section 2.3.1. The particular asymmetry required is non-axisymmetry, where the pulsar does not possess a rotational symmetry about its rotation axis. As the neutron star cools after formation, it is thought that it may form a solid crystalline crust on its surface, surrounding a core of degenerate material [10]. This crust may support a small bump or mountain, which would lead to a triaxial mass distribution, or the magnetic field of the star may deform it or cause accretion to be concentrated onto a small area of its surface. This latter possibility is especially likely in low mass X-ray binary systems, where accretion can often spin up the neutron star to periods on the order of $1\,\mathrm{ms}$.

The typical size of such a deviation from spherical symmetry is uncertain, and depends on the shear strength of the crust σ, with estimates of the allowed maximum ellipticity given by

$$\epsilon_{\mathrm{max}} = 5 \times 10^{-7} \left(\frac{\sigma}{10^{-2}} \right). \tag{1.16}$$

Estimates of the shear strength of the crust, which is a measure of the shear which can be withstood before breaking, can vary from model to model, with the largest being $\sigma_{\mathrm{max}} \approx 10^{-2} - 10^{-1}$ for a perfect crystal, but with more recent estimates being in the range $\sigma_{\mathrm{max}} \approx 10^{-4} - 5 \times 10^{-3}$ [10].

Binary Stars

Binary stars provide the most certain source of gravitational waves, and indeed it was through the observation of the decaying orbit of binary pulsar PSR B1913+16 that the first indirect measurement of gravitational waves was obtained, winning Hulse and Taylor the 1993 Nobel Prize in Physics and confirming the predictions of

General Relativity to an accuracy which with continuing observations has reached 0.2% [11, 12].

As was described in the previous section, it is the emission of gravitational waves that causes the orbit to decay, gradually shortening the period. It is expected that there is a large population of binary stars in the galaxy which are emitting gravitational waves with a range of frequencies up to 0.1 Hz. Estimates of the number of these sources based on models of the galactic population speculate that around 10^5 binaries lie in the range 1 mHz to 5 mHz alone [13].

This vast number of binaries presents a serious data analysis challenge for the proposed LISA mission, where the superimposed signals from these binaries produce an unresolvable background of gravitational waves below 1 mHz, swamping the instrumental noise curve. Above 1 mHz, these binaries begin to become individually resolvable, and their location and parameters may be estimated, although they are so dense at low frequencies that the signals will become confused. In chapter 5 I present work aimed at developing a data analysis system which can estimate these parameters where possible, and simultaneously estimate the level of the gravitational wave noise floor. Such analysis techniques will be crucial in the analysis of LISA data, which is why a series of mock data challenges are ongoing to spur development of suitable algorithms [14, 15].

1.3.3   Stochastic Sources

A stochastic gravitational wave background is postulated to exist, produced by astrophysical and cosmological sources which are not resolvable individually. There are two main candidate sources for such a background. The first is the superpo-

sition of the radiation from a vast number of binary star systems in the Milky Way which radiate at frequencies too low to be resolved in position on the sky. These sources will contribute to the data analysis challenge of the LISA mission as their combined amplitude is enough to swamp the instrumental noise curve below frequencies of approximately 1 mHz. Above this frequency the sources begin to become individually resolvable and therefore no longer constitute a true stochastic background.

A stochastic background radiation of cosmological origin is also posited to exist as a result of processes in the early universe. In an analogous fashion to the production of the cosmic microwave background radiation, there is expected to be a cosmic gravitational wave background radiation which would have been produced when the graviton decoupled from the other fields in the very early universe. Although the details of this process remain unknown, it is expected that such a decoupling would have occurred at or around the Planck epoch, 5×10^{-44} seconds after the initial singularity [16]. As such, the observation of such a relic could provide a valuable insight into the state of the universe at the earliest stages of its evolution that could be attained no other way. It has also been proposed that the process of inflation, and other phase transitions in the early universe could generate a cosmological background [17]. Different theories of cosmology and unified physics produce different predictions of the specific nature of this background, and it therefore could also be used as an observation to discriminate between them.

Although current ground-based detectors are unlikely to to observe a stochastic background, they have been able to place upper limits on the strength of such radiation, and this in turn allows a limit to be placed on Ω_{GW}, the energy density

of gravitational waves in the universe [18, 19].

1.4 Detection of Gravitational Waves

1.4.1 The Response Function of a Laser Interferometer

The interaction of the gravitational wave with an interferometric detector is anisotropic. Although it is impossible to "aim" the detector as if it were a telescope, it has a response function which is more sensitive in certain directions than in others. As the Earth rotates and orbits the Sun it carries the detector with it, sweeping the response function across the sky and modulating the amplitude of the sinusoidal signal emanating from a fixed sky position.

The response in the detector is the combination of the strains from each polarisation state, multiplied by the response to that polarisation,

$$h(t) = h_+(t)F_+(t) + h_\times(t)F_\times(t). \tag{1.17}$$

The response functions or beam pattern functions $F_+, F_\times$ are derived in [20], and have the form:

$$F_+(t) = \sin\zeta\left[a(t)\cos 2\psi + b(t)\sin 2\psi\right] \tag{1.18}$$

$$F_\times(t) = \sin\zeta\left[b(t)\cos 2\psi - a(t)\sin 2\psi\right], \tag{1.19}$$

where ζ is the angle between the arms of the detector, and the full expressions for the functions $a(t)$ and $b(t)$ which describe the rotation of the detector relative to a fixed point can be found in [20]. ψ is the angle between the polarisation axis

of the incoming gravitational wave and that of the detector, and is an unknown

parameter which must be inferred during the analysis of the data (see table 2.1).

Chapter 2

Probability Theory for Data Analysis

2.1 Bayesian Inference

When analysing gravitational wave data for a possible weak signal, it is desirable to extract the maximum possible information from the data to gain the best sensitivity. The outcome of a search for a particular type of signal is the answer to the question, *how confident am I that there is a signal present, given data* $\{B_k\}$? This is a conditional probability which we can use Bayesian Inference to express quantitatively as a value between 0 and 1. If the model of the signal has variable parameters, we can also then ask the probability density distribution as a function of the parameter space, which allows an inference to be made about these parameters, for example the amplitude. This definition of probability as a degree of reasonable confidence differs from the more common definition of a limiting frequency in repetitions of a random experiment which occurs in orthodox or fre-

quentist statistics. However, the frequentist view only gained prominence some time after the original theory of probability was developed by Laplace, who redis-covered a result published posthumously in 1763 by the Reverend Thomas Bayes [21]. It is worthwhile to briefly revisit the reasoning that led Laplace to define a probability as a degree of confidence, and a much fuller discussion of this can be found in the literature [22, 23, 24].

Laplace started with the axioms of probability, namely the sum rule

$$P(A|C) + P(\bar{A}|C) = 1, \tag{2.1}$$

and the product rule

$$P(AB|C) = P(A|BC)P(B|C). \tag{2.2}$$

In the terminology of the above equations $P(A|C)$ represents the probability of proposition A conditional on C, $\bar{A}$ is the negation of A or "not A" in the language of logic. Since any proposition must be true or false, the sum rule is an obvious requirement of the theory. $P(AB|C)$ is the probability of A *and* B being true, given C. Note that by exchanging A and B the product rule can equally well be written $P(AB|C) = P(B|AC)P(A|C)$, as the logic operation *and* is commutative.

From equation 2.2 and the commutativity relation, it is simple to derive Bayes' theorem. Consider an example where a hypothesis H is being compared with observation data d. It is also necessary to include I, representing any assumptions or prior information that may be pertinent to the problem. By simply rearranging

equation 2.2 we obtain Bayes' theorem.

$$P(H|d, I) = \frac{P(d|H, I)P(H|I)}{P(d|I)} \tag{2.3}$$

Examining equation 2.3, the meaning of each term can be distinguished. $P(H|d, I)$ is called the *posterior* probability, and is the probability of hypothesis H conditional on the data. $P(d|H, I)$ is the *likelihood* function, which is a measure of how well the data fit the hypothesis, which together with the *prior* $P(H|I)$ it allows the comparison of different hypotheses or models. $P(d|I)$ is a normalisation factor, the *marginalised likelihood* or evidence, which can be ignored when comparing different models against the same dataset, giving the relation

$$P(H|d, I) \propto P(d|H, I)P(H|I). \tag{2.4}$$

Bayes' theorem therefore allows one to assign a quantitative probability to a specific hypothesis, in light of the observations. This is of great use in data analysis where competing hypotheses are tested against observational evidence, and when comparing two competing hypotheses it is common to take the ratio of their probabilities. In this case the marginalised likelihood term cancels and equation 2.4 is a sufficient quantity to calculate for each model. The proportional form of Bayes' theorem is used extensively in the work presented here, where we are usually concerned with the relative probability of competing propositions where the hypothesis is parametrised by certain variables, and the task is to find how the probability varies as a function of them.

2.2   Parameter Estimation

Consider a measurement of a single quantity x which is parametrised to take a value in the range $x \in [0, 3)$. The measurement process is imperfect and does not yield consistent answers, so I might make multiple measurements of x and ask for the probability that x lies in a certain range of its parameter space. From the axioms of probability 2.1, 2.2 we can break down the problem thus,

$$P(0 \leq x < 1|d, I) + P(1 \leq x < 2|d, I) + P(2 \leq x < 3|d, I) = \sum_z P(x \in z|d, I) = 1,$$

since the ranges codified z are mutually exclusive and exhaustive of the parameter space. It is obvious that one can divide up the parameter range into any number of such propositions, and that in the limit of an infinite number, the sum becomes an integral

$$\int_{z=0}^{3} p(x = z|d, I)dz = 1.$$

It should be noted that the symbol p is used for the integrand to emphasise it is a probability density function, as opposed to the probabilities denoted P above. In this manner we can estimate a continuous parameter by assigning a posterior probability density function (PDF) to it, and proceeding through the use of Bayes' theorem. It is also usual to omit the dummy variable z, so Bayes' theorem as applied to estimation of a parameter μ is

$$p(\mu|d, I) = \frac{p(\mu|I)p(d|\mu, I)}{p(d|I)}. \tag{2.5}$$

If the model is a function of more than one variable, then the posterior density

function simply becomes a joint probability distribution on these variables, which may be correlated. The dimensionality of this can be reduced by the process of marginalisation to eliminate parameters that are not of interest from the result thus,

$$p(\mu|d, I) = \int_{-\infty}^{\infty} p(\mu, \theta|d, I)d\theta, \tag{2.6}$$

where θ is the parameter to be marginalised over. With these results, we are ready to proceed to analysing data from gravitational wave antennae and searching for parametrised signals.

2.3   Bayesian Inference and Gravitational Wave Data Analysis

In order to use coherent integration to raise the signal-to-noise ratio of possible signals, it is necessary to have a model of the gravitational wave signal of that source. In this section I will derive these signal models using the quadrupole formalism, which allows the definition of the likelihood function and the rest of the analysis to take place.

2.3.1   Continuous Wave Signal from a Triaxial Neutron Star

Consider a neutron star, rapidly rotating with a small equatorial ellipticity ϵ. The star has moments of inertia I_{xx}, I_{yy}, I_{zz} about three principal axes. The equatorial

ellipticity ϵ is given by

$$\epsilon = \frac{I_{xx} - I_{yy}}{I_{zz}}. \tag{2.7}$$

If the star is rotating about the z-axis with an angular frequency $\omega = 2\pi f$, gravitational waves will be emitted as the elliptical mass distribution produces a varying quadrupole moment as it rotates. By combining equation 2.7 with the quadrupole formalism of equation 1.12, the nature of the gravitational waveform can be derived [25]. Expressed as a function of ellipticity, the dimensionless gravitational wave amplitude at Earth of a pulsar at distance r is given by the expression

$$h_0 = \frac{4G}{c^4} \frac{\omega^2 I_{zz}}{r} \epsilon. \tag{2.8}$$

Which can be written with fiducial values of the parameters [26],

$$h_0 \approx 4.22 \times 10^{-24} \epsilon \left(\frac{1\,\text{kpc}}{r} \right) \left(\frac{f}{1\,\text{Hz}} \right)^2 \left(\frac{I_{zz}}{10^{38}\,\text{kgm}^2} \right). \tag{2.9}$$

The sinusoidal form of the gravitational wave is modulated in both frequency and amplitude by the relative orientation of the pulsar with respect to the detector, and by the motion of both the source and the detector as it follows the Earth in its path round the Sun. For continuous wave signals, the raw data can be demodulated to correct for the Earth's motion, with the result being the gravitational wave form as would be seen at the Solar System barycentre (SSB). This is a function of several parameters which describe the neutron star system (see figure 2.3.1) and gives a gravitational waveform

$$h(t; \mathbf{a}) = \frac{h_0}{2} F_+(t; \psi) \left(1 + \cos^2 \iota \right) \cos \Phi(t) - h_0 F_\times \cos \iota \sin \Phi(t), \tag{2.10}$$
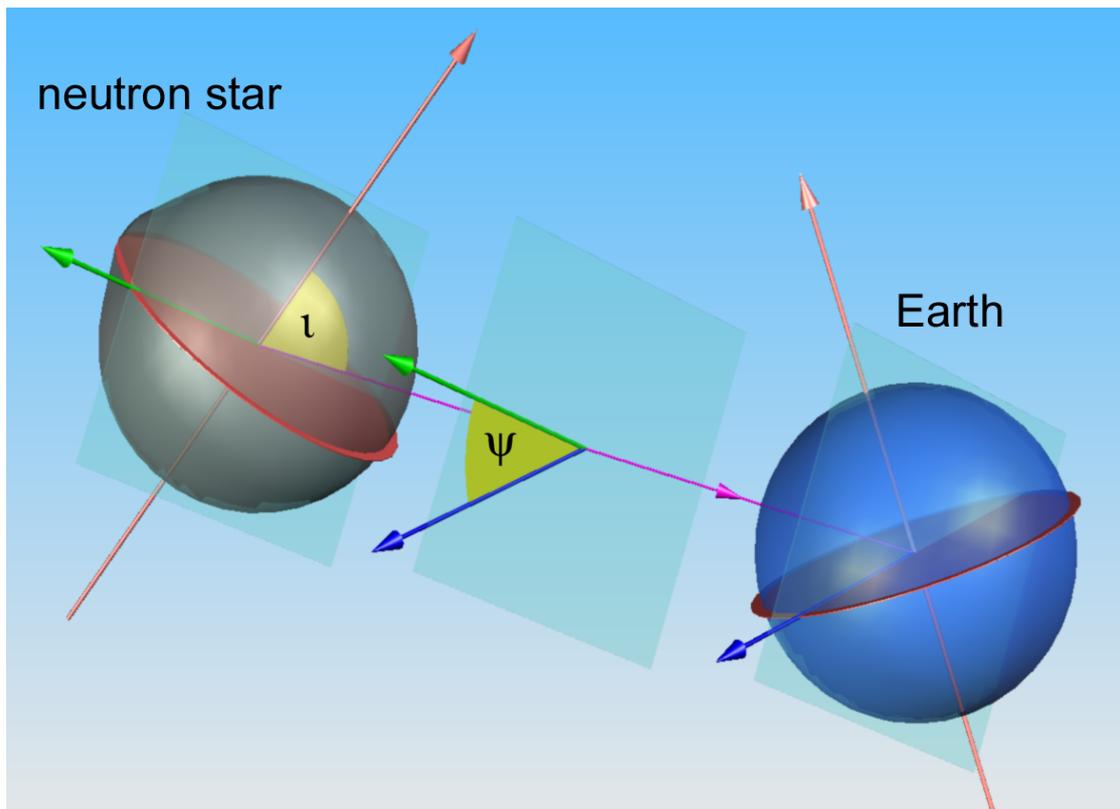
Figure 2.1: Angle parameters describing the orientation of a neutron star relative to the detector in equation 2.10. Here the Ψ parameter is defined in the opposite sense to that used in the search (ψ), such that $\psi = \frac{\pi}{4} - \Psi$. Image credit: Russell Jones

where ι is the inclination angle of the spin axis of the neutron star relative to the line-of-sight vector between the SSB and the star. F_+ and $F_\times$ are the response functions of the detector to the two gravitational wave polarisations, the magnitude of each being determined by ψ the relative polarisation angle between detector and source (see 1.4.1), and $\Phi(t)$ is the phase of the gravitational wave, which expressed as a Taylor expansion is

$$\Phi(t) = \phi_0 + 2\pi \sum_{i=0}^{\infty} \frac{t^{i+1}}{(i+1)!} \frac{d^i f}{dt^i} = \phi_0 + 2\pi \left(ft + \frac{1}{2}\dot{f}t^2 + \frac{1}{6}\ddot{f}t^3 + \dots \right). \quad (2.11)$$

Here, $t = 0$ is the epoch of the observation at which ϕ_0 is defined. It should be noted that f refers to the signal frequency, and not the rotation frequency of the pulsar. The spindown parameter $\dot{f}$ is usually very small, and further derivatives are generally unmeasurable due to timing noise [27].

When calculating the time co-ordinate used at the SSB, one must account for a variety of effects to produce a time stamp accurate enough to keep the phase of the model coherent with the phase of the signal over the period of observation. Since the time co-ordinate should be defined in a frame that is as close to co-inertial with that of the pulsar as possible, relativistic effects which are of a local cause and correctable such as the dilation of time by the presence of the Sun (Shapiro effect) and the time-varying gravitational redshift produced by the movement of the Earth (Einstein delay). If the pulsar is accelerating relative to the SSB, then this will appear as a red- or blue-shift of the source depending on whether it is accelerating away or toward the Earth respectively. In accordance with the Principle of Equivalence this is indistinguishable from a spindown or spinup of the pulsar in its own inertial frame, and so must be treated as such in the analysis.

The major contributing factor within the solar system is the Roemer delay, the travel time of the signal between the detector and the SSB, which can be up to 499 seconds (1 A.U./c) when the source, Earth and the SSB are aligned. The correct transformation from local time at the detector t_{GPS} to SSB time t is given by [28]

$$t = t_{\text{GPS}} + \Delta_{\text{Roemer}} + \Delta_{\text{Shapiro}} + \Delta_{\text{Einstein}}. \qquad (2.12)$$

With accurate tracking of these delay effects, a coherent phase model of the pulsar is maintained over the length of the observation, and this is included when calculating the downsampled signal during the first stage of the analysis. This first stage involves the multiplication of the data by a sinusoidal signal with the same frequency as the target, but with opposite phase evolution. The inclusion of these time delays in the phase evolution means that the heterodyne frequency is effectively not constant but is continually corrected to account for the particular source under consideration, as these delay effects vary with the position of the source on the sky. The process of heterodyning, which shifts a signal in frequency, is detailed further in section 4.1.1.

As the data is heterodyned and downsampled in first stage of the analysis, the variance of the noise from the detector is also calculated. Since the noise in the detector is typically not stationary over the entire observation run, it is estimated at a rate of 1/60 Hz, which is sufficient in that the noise does not vary significantly over the timescale of 1 minute. Once heterodyned, the mean of the data in each minute is calculated to yield B_k, which is the series of data points on which the algorithm will operate. In the process of downsampling the data, the variance of each sample B_k is computed from the original data points to yield σ_k, the level

of noise as estimated from a 4 Hz window around the target frequency. This is described in greater detail in section 4.1.

After the complex heterodyne, which essentially multiplies the signal by one of the opposite phase, equation 2.10 evaluated at timestamps t_k becomes [28]

$$y(t_k; \mathbf{a}) = \left(\frac{1}{4} h_0 F_+(t_k; \psi)(1 + \cos^2 \iota) - \frac{i}{2} h_0 F_\times(t_k; \psi) \cos \iota \right) \exp(i\phi(t_k, \delta f, \dot{\delta f})),$$
(2.13)

where the phase in the heterodyned signal is given by the deviation from the frequency and spindown parameters used in the heterodyne (δf and $\dot{\delta f}$ respectively) taken at epoch T_0:

$$\phi(t_k, \delta f, \dot{\delta f}) = \phi_0 + 2\pi \left(\delta f(t_k - T_0) + \frac{1}{2} \dot{\delta f}(t_k - T_0)^2 \right),$$
(2.14)

and $\mathbf{a}$ is a vector of the six parameters, $\mathbf{a} = (h_0, \cos \iota, \phi_0, \psi, \delta f, \dot{\delta f})$.

2.3.2 Likelihood function

With all these details under consideration, we are now able to write the likelihood function for a pulsar signal parametrised by the six unknown variables in table 2.1, and a vector in the 6-dimensional parameter space is denoted $\mathbf{a}$.

With the assumption that the noise follows a stationary distribution with a mean of 0 and a known variance σ_k, we may use the Gaussian distribution to model it. The set of observational data $\{B_k\}$ is therefore assumed to be composed of a signal (whose amplitude may be zero) superimposed on uncorrelated Gaussian noise. The likelihood function in equation 2.3 is then given by the product of the

Parameter	Description
h_0	Gravitational wave strain amplitude.
ψ	Polarisation angle of the source relative to the detector.
ϕ_0	Initial phase of the gravitational wave signal.
$\cos\iota$	Cosine of inclination angle of pulsar spin axis to line of sight.
f	Frequency of the gravitational wave signal ($f = 2f_{\rm rot}$).
$\dot{f}$	First time derivative of signal frequency.

Table 2.1: List of parameters of MCMC pulsar search

individual likelihoods for each data point,

$$
p(\{B_k\}|\mathbf{a}, I) = \prod_{k=1}^{N} \frac{1}{2\pi\mathscr{R}(\sigma_k)\mathscr{I}(\sigma_k)} \exp\left[-\frac{1}{2}\left|\frac{B_k - y(t_k; \mathbf{a})}{\sigma_k}\right|^2\right] \quad (2.15)
$$

$$
= \exp\left[-\frac{1}{2}\sum_{k=1}^{N}\left|\frac{B_k - y(t_k; \mathbf{a})}{\sigma_k}\right|^2\right]\prod_{k=1}^{N}\left[\frac{1}{2\pi\mathscr{R}(\sigma_k)\mathscr{I}(\sigma_k)}\right] \quad (2.16)
$$

$\mathscr{R}(\sigma_k)$ and $\mathscr{I}(\sigma_k)$ are the real and imaginary parts of the complex number σ_k respectively. When performing parameter estimation over a fixed data set of known σ_k, the product term is irrelevant and we have

$$
p(\{B_k\}|\mathbf{a}, I) \propto \exp\left[-\frac{1}{2}\sum_{k=1}^{N}\left|\frac{B_k - y(t_k; \mathbf{a})}{\sigma_k}\right|^2\right]. \quad (2.17)
$$

This function, when multiplied by the prior, gives the posterior probability density function $p(\mathbf{a}|B_k, I)$ of a gravitational wave of parameters $\mathbf{a}$ being present in the data. This is defined on the six-dimensional parameter space of $\mathbf{a}$, which has the ranges shown in table 2.2. These ranges are a constraint on the prior probability distribution, as any possibility outside of them is effectively assigned a probability of zero.

Parameter	min.	max.
h_0	0	1×10^{-19}
ψ	$-\frac{\pi}{4}$	$\frac{\pi}{4}$
ϕ_0	0	2π
$\cos\iota$	-1	1
δf	$-\frac{1}{120}$ Hz	$\frac{1}{120}$ Hz
$\dot{\delta f}$	$-1 \times 10^{-9}\,\mathrm{Hz\,s^{-1}}$	$1 \times 10^{-9}\,\mathrm{Hz\,s^{-1}}$

Table 2.2: The range of each parameter describing the pulsar waveform after heterodyne 2.13.

By examining equations 2.13 and 1.18, it is clear why the range of the angle parameter ψ covers only the interval $[-\frac{\pi}{4}, \frac{\pi}{4}]$. The factor of two produces a copy of the function under the transformation $\psi \to \psi + \pi$, rendering half the the range $[-\pi, \pi]$ unnecessary. The range is further reduced by one half because the transformation $\psi \to \psi + \frac{\pi}{2}$ is equivalent to $\phi_0 \to \phi_0 + \pi$. By allowing ϕ_0 to vary over the full 2π range, we may therefore limit the range of ψ to that shown in the table.

2.3.3 Priors on Parameters

In order to define the PDF, the prior probability distribution must also be known or assigned based on assumptions or relevant information that might be known. In the case of the pulsar orientation parameters which are unknown, the prior should represent an equal probability of the pulsar spin axis pointing in any direction on the sphere surrounding it, i.e. a uniform prior on area. The element of area as expressed in polar co-ordinates ψ, ι is $dA = \sin\iota\, d\iota\, d\psi = d\cos\iota\, d\psi$, so the prior should be uniform on the parameters $\cos\iota$ and ψ, which allows convenient use of $\cos\iota$ as a parameter rather than ι. The phase of the signal ϕ_0 is also unknown,

so the non-informative uniform prior over the range is used here also. The dimensionless amplitude parameter h_0 was assigned a prior which was uniform over the range $(0, 10^{-19})$. It may be argued that it should have an "scale prior" of the form $p(h_0|I) \propto \frac{1}{h_0}$ [22], however this would lead to a distribution which is formally unnormalisable, as it diverges at $h_0 = 0$. The use of this prior also presents difficulties for Markov Chain Monte Carlo estimation, as introduced in 2.4, since the chain tends to become trapped in the area close to the origin where the diverging prior overwhelms the likelihood, rather than exploring the full parameter space. In practice, any prior which does not assign probability zero to the likelihood mode will return very similar results, as the product of the individual data likelihoods overwhelm the prior when their number is large, as is in this case with typically tens of thousands of points.

The prior distributions on frequency and spindown were adjustable, taking two possible forms. The first and simpler of these were simply uniform priors over the entire range of frequencies permitted by the Nyquist theorem. This was used in the broadband search where the frequency of the signal was considered to be essentially unknown. The alternative prior was a normal distribution that could be centred on a target frequency to perform a more detailed search of a narrow range of frequencies.

2.4   The Markov Chain Monte Carlo Algorithm

Having defined the posterior PDF for a pulsar signal in the data, the problem of searching for a signal becomes one of examining this distribution. If a signal is present, it will appear as a strong peak at some particular values of the parameters;

the maximum value and width of the peak are determined by the signal to noise ratio of the source. For a strong signal the probability distribution surrounding the peak is well approximated by a multivariate normal distribution in the six parameters, however as we shall see certain combinations of parameters are highly correlated, and the width of the distribution in frequency is exceedingly narrow compared to the range of the search. In previous work on targeted pulsars, where the frequency and spindown are known from radio observations, an exhaustive search of the 4-dimensional parameter space was possible by placing a grid over the 4 parameters and evaluating the likelihood. If no maximum was found indicating a signal, this could then be marginalised over the ψ, ϕ_0 and $\cos \iota$ parameters to set an upper limit, as described in [28, 27, 26].

When the frequency and spindown parameters are introduced into the search it becomes no longer practical to search the parameter space exhaustively, based on a grid. The reason for this can be seen if we consider the width of the posterior mode in frequency and spindown. For a signal which is not super-resolved, i.e. one at low SNR, the width of the main mode in frequency will be, at most, the width of one frequency bin given by $\Delta f = T_{\mathrm{obs}}^{-1}$, and in spindown $\Delta \dot{f} = T_{\mathrm{obs}}^{-2}$.

Therefore the approximate number of grid points that would be necessary for an exhaustive search of our range in frequency and spindown can be approximated as the product of these two numbers, with the assumption that a rectangular grid is used,

$$N \approx (f_{max} - f_{min}) \times \left(\delta \dot{f}_{max} - \delta \dot{f}_{min} \right) \times T_{\mathrm{obs}}^3,$$

which for a run of length 30 days ($= 2\,592\,000$ seconds) and parameter ranges as specified in table 2.2 gives $N \approx 5.8 \times 10^8$. This must then be evaluated for

each combination of the other four parameters. While there may be some savings to be made in accounting for correlations between frequency and spindown, this rough estimation shows that performing an exhaustive time-domain search of this magnitude comes at a very great computational cost.

One possible way around this problem is to use a different means of sampling the posterior PDF, such that not every combination of parameters has to be tested. One way of doing this is the technique called Markov Chain Monte Carlo, which has gained popularity amongst the physics and astronomy community with the advent of powerful computers in recent years. Here I will describe how it is implemented, and in the rest of this thesis study its application to problems of gravitational wave data analysis.

2.4.1 Markov Chains

A Markov chain is defined as a collection of samples drawn from some range of possibilities in sequence, such that the $(n + 1)$-th sample is drawn from a distribution $p(x_{n+1}|x_n)$ dependent only on the current sample x_n and none of the previous states of the chain. The space from which the samples are drawn is known as the state space of the chain. The probability of moving from one state i to another j in one step is the transition probability p_{ij}; if this is unchanging during all steps of the chain, i.e. $p(x_{n+1}|x_n) = p(x|x_{n+1})$, it is said to be time-homogeneous. Furthermore, if the chain is *aperiodic*, then it will not indefinitely oscillate between states in a periodic manner, but is free to explore the full range of states. If the state space possesses the additional property of having a non-zero transition probability between every pair of states, it has a unique stationary distribution π_i, which is

normalised such that $\sum_i \pi_i = 1$. As the number of samples in the chain tends to infinity, it can be shown that the distribution of samples among the states converges to the stationary distribution. For a good introduction to these properties of Markov Chains and their applications in inference, see Gamerman 1997 [29].

If the conditions above are met, this process of convergence will occur regardless of where in the state space the chain is started. Thus the stationary distribution of the Markov chain can be efficiently estimated by examining the density distribution of samples. The process is valid for both discrete and continuous state spaces, where in the latter case a histogram can provide the means of displaying the distribution.

2.4.2 The Metropolis-Hastings Algorithm

The power of MCMC as applied to a sampling problem comes from matching the stationary distribution of the Markov chain to an arbitrary distribution which is difficult to evaluate by other means. The method of doing this was first described by Metropolis *et al* [30], and expanded upon in [31] as the *Metropolis-Hastings algorithm*. In our application to sampling the probability distribution $p(\mathbf{a}|B_k, I)$, the state space of the chain is the parameter space of the PDF, and each sample $\mathbf{a_n}$ is a vector in this parameter space.

The basic Metropolis-Hastings algorithm is very simple to implement. The chain is initialised by choosing a first sample x_0. This may be done at random or by making an informed guess, as the distribution of samples in a converged Markov chain is independent of the initial state. The algorithm then consists of two steps which are iterated until sufficient samples have been obtained:

1. A proposed sample $\mathbf{Y}$ is drawn from a *proposal distribution* $q(\mathbf{a}_{n+1}|\mathbf{a}_n)$ as in the standard Markov process. The proposal distribution is denoted q to distingush it from the probability distributions dependent on the data, which are labelled p.

2. Calculate the Metropolis ratio r,

$$r = \frac{p(\mathbf{Y}|\{d\}, I)}{p(\mathbf{a}_n|\{d\}, I)} \frac{q(\mathbf{a}_n|\mathbf{Y})}{q(\mathbf{Y}|\mathbf{a}_n)} \tag{2.18}$$

The proposal is either accepted or rejected probabilistically, with the acceptance probability given by

$$\alpha(\mathbf{a}_n, \mathbf{Y}) = \min(1, r). \tag{2.19}$$

This is done by drawing a random number from the uniform distribution $U \sim \mathrm{unf}(0, 1)$ and comparing with r:

   (a) *IF* $(U > r)$ Reject the proposal, and count the current state again. $\mathbf{a}_{n+1} = \mathbf{a}_n$.

   (b) *ELSE* Accept the proposal and append it to the chain. $\mathbf{a}_{n+1} = \mathbf{Y}$.

   Return to step 1, incrementing n.

In the original Metropolis algorithm, a symmetric proposal distribution was used with $q(\mathbf{Y}|\mathbf{a}_n) = q(\mathbf{a}_n|\mathbf{Y})$, and these terms cancel in the ratio. To prove that this procedure generates samples from the target distribution $p(\mathbf{a}_n|\{d\}, I)$, consider the joint PDF for two consecutive samples $p(\mathbf{a}_{n+1}, \mathbf{a}_n|\{d\}, I) = p(\mathbf{a}_{n+1}|\mathbf{a}_n)p(\mathbf{a}_n|\{d\}, I)$.

As is shown in [32], if the conditions in section 2.4.1 are met, the chain obeys the *principle of detailed balance*, $p(\mathbf{a}_{n+1}|\mathbf{a}_n)p(\mathbf{a}_n|\{d\}, I) = p(\mathbf{a}_n|\mathbf{a}_{n+1})p(\mathbf{a}_{n+1}|\{d\}, I)$. Using this equation and marginalising over the range of possible states of $\mathbf{a}_n$, one can obtain the sampling distribution of $\mathbf{a}_{n+1}$:

$$
\begin{aligned}
\int p(\mathbf{a}_{n+1}|\mathbf{a}_n)p(\mathbf{a}_n|\{d\}, I)\mathbf{da}_n &= \int p(\mathbf{a}_n|\mathbf{a}_{n+1})p(\mathbf{a}_{n+1}|\{d\}, I)\mathbf{da}_n \\
&= p(\mathbf{a}_{n+1}|\{d\}, I) \int p(\mathbf{a}_n|\mathbf{a}_{n+1})\mathbf{da}_n \\
&= p(\mathbf{a}_{n+1}|\{d\}, I)
\end{aligned}
$$

Therefore, by recursion, if $\mathbf{a}_0$ is a sample in the stationary distribution, each subsequent member of the Markov chain is also drawn from this distribution.

2.5 Optimisation of Markov Chain Monte Carlo

Having described above the basic Metropolis-Hastings algorithm, we are in a position to ask how it may be improved. Although the MCMC chain may converge on the signal given sufficient time, it is desirable that this happens as quickly as possible. A variety of modifications to the standard Metropolis-Hastings algorithm have been developed to speed up the convergence of the chain and improve its exploration of the parameter space. I will here describe those that have been incorporated into the search algorithm for targeted pulsars and the reasons for doing so.

The chief obstacle to speedy convergence of the Markov chain is the highly intricate structure of the likelihood surface which it is sampling. It is inevitable that in random noise there will appear features that imitate, to a greater or lesser
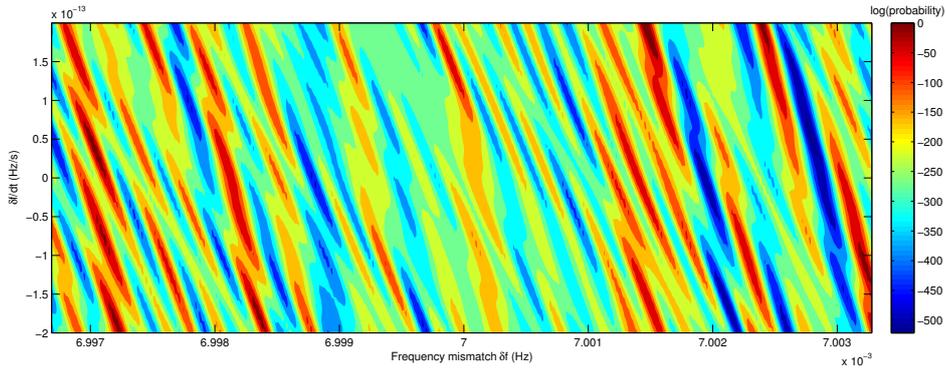
Figure 2.2: A small section of a slice through the posterior PDF in the $(\delta f, \dot{\delta f})$ plane for 63,960 s of noisy data with no signal. In order to efficiently sample this multi-modal distribution the Markov chain must be able to step between local modes to find the most probable. Note the correlation between the parameters δf and $\dot{\delta f}$ which is removed by a reparametrisation, as described in section 2.5.4.

extent, the model that one is searching for. This causes features to appear in the joint posterior PDF; local maxima that may trap the Markov chain as it performs its random walk through the space, as shown in figure 2.2. These features are independent of any true signal $s(t, a_T)$ that may be present in the data, since the data is a linear combination of the signal and noise;

$$B_k = s(t_k, \mathbf{a_T}) + N(0, \sigma^2). \tag{2.20}$$

In order to find the global maximum representing a signal (if any is present), the chain must explore between these local maxima efficiently. Evaluating such multi-modal likelihood surfaces poses a problem, as the proposal distribution must be chosen carefully to allow jumps between modes as well as sampling within a mode. We used an adaptive proposal distribution which took the form of a multivariate Gaussian distribution centred on the current state.

Three techniques were used to enhance the efficiency of the algorithm, simulated annealing (burn-in), delayed rejections and reparametrisation of the posterior PDF; which are documented below, and in [33]. Of these three, only the introduction of a burn-in period affects the Markovian property of the chain, and so samples from the burn-in period are not used to calculate the posterior.

2.5.1 Simulated Annealing

Simulated annealing, or simulated tempering, is a procedure whereby the chain's exploration of the state space is accelerated by modifying the Metropolis ratio 2.18. By introducing an inverse temperature β, the acceptance ratio of the chain is increased thus

$$r = \left(\frac{p(\mathbf{Y}|\{d\}, I)}{p(\mathbf{a}_n|\{d\}, I)} \frac{q(\mathbf{a}_n|\mathbf{Y})}{q(\mathbf{Y}|\mathbf{a}_n)} \right)^{\beta}. \tag{2.21}$$

During the annealing schedule run, the value of β is gradually increased from a starting value β_0 on an exponential curve, until it reaches the value 1 after N_b iterations, following the rule

$$\beta(N) = \begin{cases} \beta_0 \exp\left(\frac{N}{N_b} \log\left(\frac{1}{\beta_0} \right) \right) & N < N_b \\ 1 & N \geq N_b \end{cases} \tag{2.22}$$

During this period, known as the burn-in, the resulting samples are not drawn from the target distribution and so cannot be included in the final Markov chain. As shown in figure 2.3, the acceptance probability is initially inflated so as to allow the chain to make unlikely transitions and move between maxima. As the temperature decreases, the transition probabilities converge to their normal values, with the lower probabilities decreasing faster. The chain is therefore concentrated
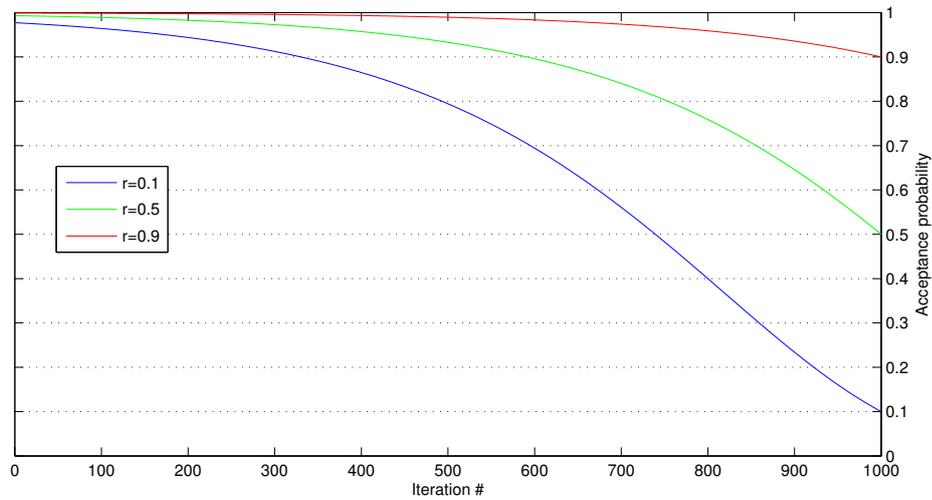
Figure 2.3: The modified acceptance probabilities for three different transitions during a burn-in phase, with $\beta_0 = 0.01$ and length $N_b = 1000$.

in the regions of high probability as transitions to less probable areas become increasingly unlikely. In this manner the chain may move between local maxima with increased odds of finding the global maximum.

The value of the burn-in parameter β_0 was chosen to be 0.01 for the pulsar search code. The length of burn-in varied at different stages in the development of the algorithm, as a long burn-in requires proportionately more processing time to execute. A default value of 1,000,000 iterations was decided upon based on experience using the algorithm to search in a real datastream of length 63 960 minutes. This experience was primarily gained by performing a large number of trial runs, and deciding on an acceptable trade-off between speed and sensitivity (see section 4.3).

2.5.2 Delayed Rejection

When a Markov chain is sampling from the mode of a PDF, it should take steps which have a characteristic length scale less than the width of the distribution being sampled. However, during the exploratory phase, the chain should make large random jumps in order to fully explore the range of possibilities. To balance the need to make small and large jumps, we have implemented a delayed rejection algorithm, as originally described in [34]. The use of delayed rejection allows information from previous iterations to be used in making the next proposal, while maintaining the stationarity condition required for a Markov Chain.

The Delayed Rejection algorithm is initially similar to the standard Metropolis-Hastings routine described in section 2.4.2. At stage one a proposal $\mathbf{Y}$ is generated as before, by drawing a sample from a multivariate normal distribution $q_1(\mathbf{a}_n)$ centred on the current position of the chain $\mathbf{a}_n$. However, if $\mathbf{Y}$ is rejected, instead of counting the state $\mathbf{a}_n$ again as in the standard algorithm, the second stage of the algorithm is entered. A second proposal $\mathbf{Z}$ is drawn from another multivariate normal distribution $q_2(\mathbf{a}_n)$ with a different covariance matrix. This allows the algorithm to make initially bold proposals to explore the space, but if they are rejected to sample the local distribution. The choice of proposal distributions will be discussed later in 2.5.3. In order to preserve the principle of detailed balance, the transition probability must be the same for a forward transition $\mathbf{a}_n \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}$ as for the reverse process $\mathbf{Z} \rightarrow \mathbf{Y} \rightarrow \mathbf{a}_n$. Therefore the appropriate reverse transition probability to use in the ratio must be calculated as if a stage one proposal $\mathbf{Z} \rightarrow \mathbf{Y}$ was rejected before the stage two proposal $\mathbf{Z} \rightarrow \mathbf{a}_n$ was selected. If the acceptance probability of the stage one transition (given by the original Metropolis ratio) is

denoted α_1, its rejection probability is then $1 - \alpha_1$ and the joint probability of making the second stage proposal $\mathbf{a}_n \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}$ is given

$$P(\mathbf{Z}, \mathbf{Y}|\mathbf{a}_n) = q_1(\mathbf{Y}|\mathbf{a_n}) \left[1 - \alpha_1(\mathbf{Y}|\mathbf{a_n})\right] q_2(\mathbf{Z}|\mathbf{a}_n, \mathbf{Y}) \tag{2.23}$$

and the reverse transition $\mathbf{Z} \rightarrow \mathbf{Y} \rightarrow \mathbf{a}_n$ probability is

$$P(\mathbf{a}_n, \mathbf{Y}|\mathbf{Z}) = q_1(\mathbf{Y}|\mathbf{Z}) \left[1 - \alpha_1(\mathbf{Y}|\mathbf{Z})\right] q_2(\mathbf{a}_n|\mathbf{Z}, \mathbf{Y}). \tag{2.24}$$

These are substituted into the Metropolis ratio where before only the proposal distribution q_1 had to be included. We therefore have a second stage acceptance ratio

$$\alpha_2(\mathbf{Z}|\mathbf{a}_n) = \min\left(1, \frac{p(\mathbf{Z}|I)p(\{d\}|\mathbf{Z}, I)}{p(\mathbf{a}_n|I)p(\{d\}|\mathbf{a}_n, I)} \frac{q_1(\mathbf{Y}|\mathbf{Z})\left[1 - \alpha_1(\mathbf{Y}|\mathbf{Z})\right]q_2(\mathbf{a}_n|\mathbf{Z}, \mathbf{Y})}{q_1(\mathbf{Y}|\mathbf{a_n})\left[1 - \alpha_1(\mathbf{Y}|\mathbf{a_n})\right]q_2(\mathbf{Z}|\mathbf{a}_n, \mathbf{Y})}\right).$$
$$\tag{2.25}$$

The transition process is depicted in figure 2.4, where the initial proposal distribution has a large variance which corresponds to an attempt to move between local maxima of the multi-modal distribution. This step is rejected, leading to a second proposal based on a distribution which matches the scale of the local mode in order to generate a valid sample from the mode in which the chain is situated. Through this process the chain will explore the space while large-scale jumps are still acceptable, but when the chain finds an area of high probability and stage one proposals fail, the acceptance rate remains high thanks to the second stage.
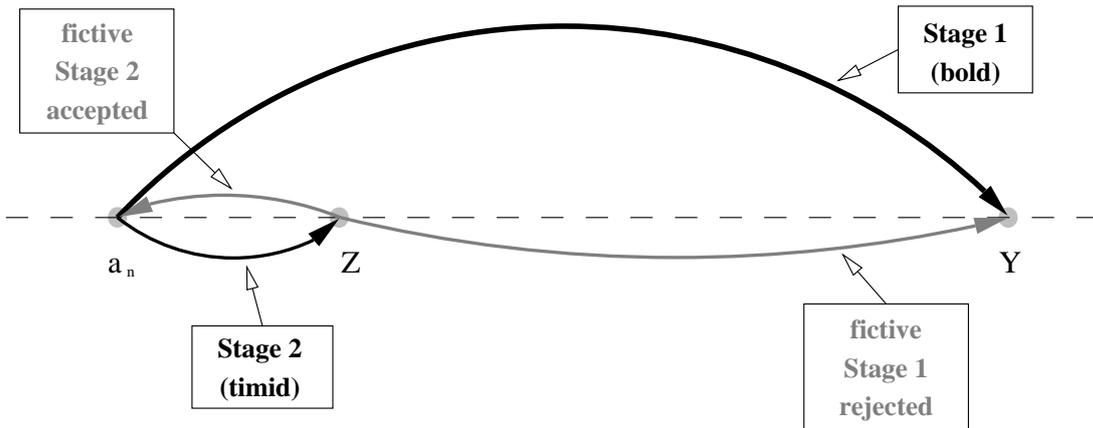
Figure 2.4: In black, the two-stage delayed rejection algorithm initially makes a bold proposal, which if rejected is followed by a more conservative second stage proposal. In grey is shown the reverse transition, as in equation 2.24, where a fictive step $\mathbf{Z} \rightarrow \mathbf{Y}$ is rejected and followed by the reverse second stage $\mathbf{Z} \rightarrow \mathbf{a}_n$. Figure from Umstätter *et al* [33].

2.5.3 Proposal Distributions

As mentioned above, the choice of proposal distribution is an important factor in ensuring the successful convergence of a Markov Chain. In the context of a multi-modal posterior, such as we have in the case of searching for a pulsar in noisy data, it is important that proposals are made both on the scale of the entire parameter space and on the scale of the local modes. Failure to make large moves means the PDF is inadequately explored, but failure to make small moves will lead to a very low acceptance rate and poor mixing of the chain. The delayed rejection algorithm requires proposals to be made on two scales in an attempt to provide a mixture of the two. These two scales are specified by manually chosing numbers, as they may need to be varied depending on the data that is being analysed. If a datastream contains a loud signal with high signal to noise ratio (SNR) the scale of the local mode will be smaller than that from a quiet signal. In addition to this

flexibility, there is also a probabilistic element to the choice of the scale used for the proposal distribution at each iteration which allows proposals to be made at scales intermediate to the specified values. This is accomplished by calculating a variance at random from an exponential curve between the two scales.

For each parameter, the two scale factors $\sigma_{\text{low}}, \sigma_{\text{high}}$ are combined to create the chosen variance $\sigma = \sigma_{\text{low}}^{a} \times \sigma_{\text{high}}^{1-a}$, where a is a random parameter in the range $(0, 1)$ which weights each possibility. The distribution from which a is drawn varies between stages so as to favour either σ_{low} or σ_{high}.

During stage one proposals, a is drawn from a beta distribution $a \sim B(2, 1)$, whereas during stage two $a \sim B(1, 2)$. The parameters of the beta distribution determine whether the density of a is higher toward the lower or higher end of the range. The beta distribution probability density function on variable x is given by the formula in terms of shape parameters α and β, and the gamma function $\Gamma()$ as,

$$B(x, \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1} \tag{2.26}$$

This is a simple means of ensuring that a variety of step sizes are proposed, as the proposal distribution has no dependence on the parameter values; instead it is tuned by hand for the particular application. The major contributory factor to this tuning is the signal to noise ratio in the data - as the LIGO interferometers improved the noise level has dropped. In the absence of a signal, the width of the h_0 posterior is determined by the noise, and therefore smaller steps in the amplitude parameter are required.
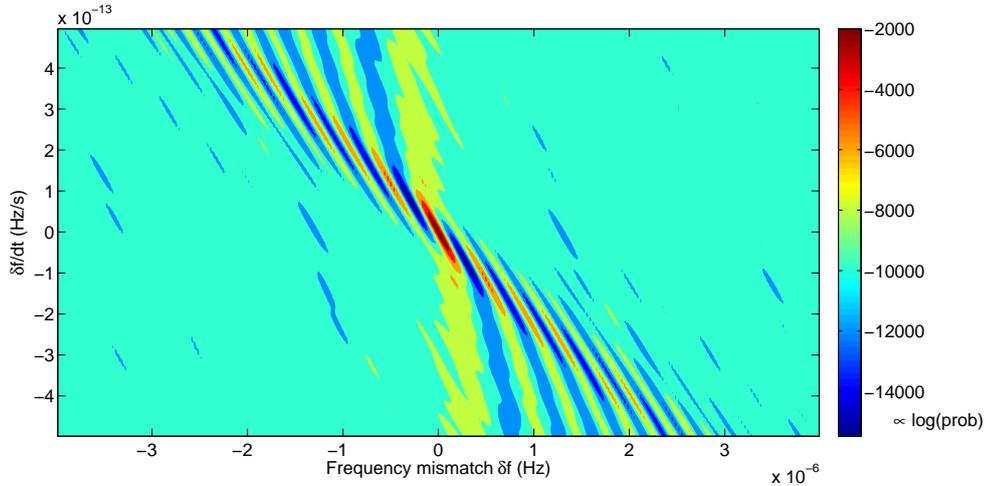
Figure 2.5: As in fig 2.2, a small section of the posterior PDF in the $(\delta f, \dot{\delta f})$ plane for $T_{\mathrm{obs}} = 63,960\,\mathrm{s}$ of noisy data with a signal injected at $\delta f = 0, \dot{\delta f} = 0$ and amplitude $h_0 = 1 \times 10^{-23}$. This clearly shows the maximum probability at the injected point, with probability side-lobes (of width $1/T_{\mathrm{obs}}$ Hz when marginalised onto the δf axis) extending in the plane. If the Markov Chain falls into one of these side lobes it may step upwards towards the true maximum width greater speed if the appropriate reparametrisation is made to align the proposal distribution with the modes.

2.5.4 Reparametrisation

Correlation of Parameters

With the above optimisations in place, the MCMC search algorithm has an improved efficiency in finding the global maximum of probability. However, trial runs showed that once converged on the peak, the acceptance of proposals was impeded by the correlation of certain parameters. In particular, as can be seen from figures 2.5 and 2.6, there is a strong correlation between the parameters δf and $\dot{\delta f}$ and between h_0 and $\cos \iota$.

Using a multivariate normal distribution with no off-diagonal elements is a poor
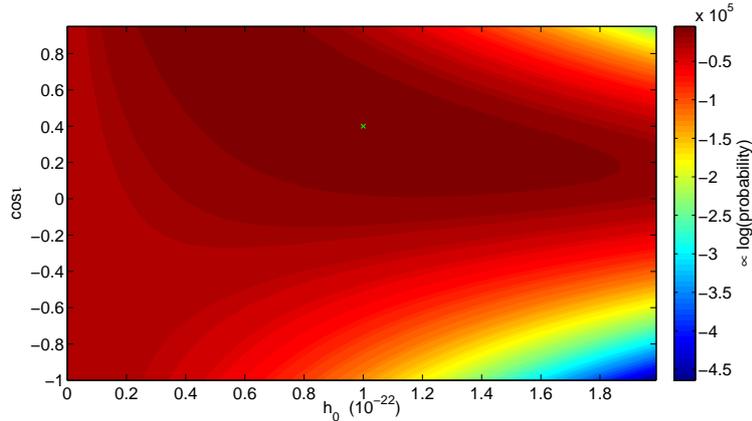
Figure 2.6: A section of the posterior PDF in the $(h_0, \cos\iota)$ plane for the same signal as in figure 2.5. In this plot the structure of the PDF in these parameters is revealed to be non-Gaussian and correlated, which necessitates the reparametrisation of these parameters for speedy mixing of the chain. The true signal was injected at the point marked $\times$.

choice of proposal distribution for distributions which are so correlated. To solve this problem one could insert non-diagonal elements into the covariance matrix of the proposal distribution so that it would more closely match that of the target distribution, or cast the likelihood function into new variables.

Changing Variables

As the correlation between the h_0 and $\cos\iota$ parameters is not constant throughout the parameter space, it was decided that a reparametrisation of the correlated variables would be the preferable solution to this problem. An overview of the procedure is given in [33] which is explained in more detail here.

The subroutine of the program which generates proposals was therefore altered, such that instead of drawing from a multivariate normal distribution of the six variables in 2.1, it would change variables to the new uncorrelated parameters,

draw the proposal and convert back into the original variables before passing it back to the main loop of the MCMC code. In this way the alterations to the program were localised to only one function.

Throughout the period of observation, the instantaneous signal frequency after heterodyning varies as in the Taylor expansion 2.14 between the values

$$f_{\text{start}} = \delta f + \frac{1}{2}\dot{\delta f} t_{\text{start}} \tag{2.27}$$

and

$$f_{\text{end}} = \delta f + \frac{1}{2}\dot{\delta f} t_{\text{end}}, \tag{2.28}$$

where t_{start} and t_{end} are the start and end times of the observation. These are taken as the new parameters, and steps are then proposed which are either highly correlated so as to vary both f_{start} and f_{end} together (corresponding to a uniform change in frequency), or vary them with no correlation so as to change $\dot{\delta f}$ indirectly. These two options are chosen between at random with equal probabilities by drawing from the uniform distribution $U \sim \text{Unf}(0, 1)$. Since the scales of f_{start} and f_{end} are the same, they may have the same variances, and the correlation sub-matrix for these parameters is given

$$C_{ij}^f = \begin{cases} \begin{bmatrix} \sigma_f^2 & 0.999\sigma_f^2 \\ 0.999\sigma_f^2 & \sigma_f^2 \end{bmatrix} & U < 0.5 \\ \begin{bmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_f^2 \end{bmatrix} & U \geq 0.5 \end{cases} \tag{2.29}$$

The 0.999 values differ from unity to make the matrix invertible.

The new values of the δf and $\dot{\delta f}$ parameters are then recovered by performing the inverse transformation

$$\dot{\delta f} = 2\frac{f_{\text{end}} - f_{\text{start}}}{t_{\text{end}} - t_{\text{start}}} \tag{2.30}$$

and

$$\delta f = f_{\text{start}} - \frac{1}{2}\dot{\delta f}_{\text{start}}. \tag{2.31}$$

The reparametrisation of the h_0 and $\cos\iota$ parameters proceeds according to the non-linear transformation

$$a_1 = \frac{1}{4}(1 + h_0 \cos^2\iota) \tag{2.32}$$

$$a_2 = \frac{1}{2}h_0 \cos\iota. \tag{2.33}$$

These two parameters correspond to the amplitudes of the real and imaginary parts of equation 2.13. A similar procedure as above is applied in choosing the covariance sub-matrix for these parameters, except that an anticorrelation applies:

$$C_{ij}^a = \begin{cases} \begin{bmatrix} \sigma_a^2 & -0.999\sigma_a^2 \\ -0.999\sigma_a^2 & \sigma_a^2 \end{bmatrix} & U < 0.5 \\ \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_a^2 \end{bmatrix} & U \geq 0.5 \end{cases} \tag{2.34}$$

Where an uncorrelated change corresponds to an overall variation of amplitude of the signal, and an anticorrelated change to a shifting in power between the $+$ and $\times$ polarisation states of the gravitational wave as would be produced by a change in inclination angle ι.

The reverse transformation in this case is found by inverting the equations to

find

$$h_0 = 2 \left(a_1 + \sqrt{a_1^2 - a_2^2} \right) \tag{2.35}$$

$$\cos \iota = 2 \frac{a_2}{h_0}. \tag{2.36}$$

Since the transformation of these parameters is non-linear, the probability density in a_1, a_2 is not equal to that in $h_0, \cos \iota$. This may be thought of as a distortion of the joint prior probability distribution which must be corrected to ensure that samples are being drawn from the correct target with uniform priors on h_0 and $\cos \iota$. As derived in [33], the prior PDF in the original parameters may be written

$$p(h_0, \cos \iota | I) = \begin{cases} 2h_{\max}^{-1} & 0 \leq h_0 \leq h_{\max}, \\ 0 & \text{otherwise} \end{cases} \tag{2.37}$$

which in the new parameters is,

$$p(a_1, a_2 | I) = \begin{cases} 2h_{\max}^{-1} |J|, & |a_2| < a_1 < \frac{4a_2^2 + h_{\max}^2}{4h_{\max}} \leq \frac{h_{\max}}{2} \\ 0, & \text{otherwise} \end{cases} \tag{2.38}$$

where $|J|$ is the Jacobian

$$|J| = \frac{2}{\sqrt{a_1^2 - a_2^2}}. \tag{2.39}$$

In this reparametrisation the limits of the distribution, as expressed in the new parameters, have been chosen so as to ensure that upon returning to the original parameters, h_0 and $\cos \iota$ remain real.

2.5.5 Behaviour at Edges of Parameter Space

From time to time, when the chain is close to the edge of the allowed prior range of parameter space a proposal will be made which attempts to move beyond this range. The model contains parameters which are both cyclical in the case of the angle parameters ψ, ϕ_0; and non-cyclical in the case of the frequency, spindown and $\cos\iota$. The implementation of the algorithm distinguishes between these two cases when a step is made beyond the boundary.

For the latter case, the proper behaviour of a Markov chain upon proposing such a step is to treat it as any other proposal and evaluate the probability density at this point. Since the prior here is by definition equal to zero, the step is always rejected, causing the current sample to be repeated in the chain. It is important that such proposals are allowed to be processed by the algorithm, otherwise there will be a bias near the edges reducing the sample density there.

For the parameters ϕ_0 and ψ, implementing the same range checking would be an acceptable solution, but a cursory inspection of the model equation 2.13 makes it obvious that the likelihood is the same at $\phi_0 = 0$ and $\phi_0 = 2\pi$. Therefore, when a step is made beyond the edges of the parameter space in this variable, it has a natural mapping back inside the allowed range, implemented in software within the proposal routine as $\phi_0 \rightarrow \phi_0 \,(\mathrm{mod}\, 2\pi)$.

Similarly, if a proposal is made beyond the limits of ψ, such that $\psi > \frac{\pi}{4}$, this may be mapped back into the parameter space with the transformation $\psi \rightarrow \psi - \frac{\pi}{2}$, $\phi_0 \rightarrow \phi_0 + \pi$. Likewise if $\psi < \frac{\pi}{4}$, the transformation $\psi \rightarrow \psi + \frac{\pi}{2}$, $\phi_0 \rightarrow \phi_0 + \pi$ applies.

In the case of the dimensionless amplitude h_0, the transformations $h_0 \rightarrow -h_0$

and $\phi_0 \rightarrow \phi_0 + \pi$ are equivalent, so an attempted jump in h_0 below zero is mapped $-h_0 \rightarrow h_0$, $\phi_0 \rightarrow \phi_0 + \pi$.

These mappings increase the acceptance ratio of the chain if it is close to the edge of the parameter space, and therefore the efficiency of the sampling. For example if there were a probability peak close to $\phi_0 = 0$ with a width such that there was a high probability density near both $\phi_0 = 0$ and $\phi_0 = 2\pi$, allowing steps across the boundary would effectively reduce this bimodal maximum to a unimodal one and aid convergence.

2.6 Concluding Remarks

With the information presented above on the nature of the problem of searching for gravitational radiation from an isolated pulsar of uncertain frequency, and on the methods of Bayesian inference, I can now present an analysis pipeline implementing a search for this radiation and setting upper limits in its absence. Having described above the core MCMC data analysis algorithm, developed by myself and others, in the next chapter I will examine its performance on simulated and real data from the LIGO Hanford Interferometer, particularly in a search for a candidate pulsar in the remnant of Supernova 1987A.

Chapter 3

Testing of the MCMC Algorithm for Neutron Star Searches on Artificial Data

In this chapter I shall first describe the testing of the MCMC algorithm on simulated data, showing the recovery of the injected signal parameters and investigating the limits of detectability. I shall then proceed to describe the search pipeline, which enables the algorithm to be used in the search for continuous gravitational waves in data from LIGO Hanford interferometer. This pipeline is also designed to allow the setting of upper limits on amplitude, in the event that no detection is made, via a Monte Carlo injection procedure. Finally I shall use the pipeline to set upper limits on the amplitude of gravitational radiation emanating from the putative pulsar in the remnant of Supernova 1987A [35], searching over a frequency range of $4\,\mathrm{Hz}$ and a spindown range of $2 \times 10^{-10}\,\mathrm{Hz\,s^{-1}}$.

3.1 Performance of Algorithm on Simulated Data

Having outlined the MCMC algorithm used in chapter 2, it is necessary to test the specific implementation of this in software. The reasons for this are both to debug the algorithm and eliminate errors in coding, and to evaluate its performance in a realistic search for a signal. Here I shall use the algorithm to perform parameter estimation on data where the parameters are known in advance, which will highlight both the capabilities and limitations of this approach.

To check the implementation of the MCMC algorithm, it was initially tested on artificially generated white noise, so as to ensure that no spectral lines, glitches or other artifacts were present that may otherwise interfere with the analysis. In order to accomplish this, a program was developed to inject an arbitrary pulsar signal of any desired parameters into the simulated data files. This was a small piece of code which read in the data and desired signal parameters, calculated the pulsar gravitational wave form at each timestamp of the data and output the sum of the data and the signal. The waveform was calculated according to the post-heterodyne equation 2.13, where the amplitude response functions were calculated using the LIGO Algorithm Library (LAL [36]) function `LALComputeDetAMResponse()`. The documentation for this software package can be found online at [37]. As this function requires the user to specify the interferometer for which the response function is to be generated, I chose the LIGO Hanford Observatory 4km (H1). The LAL functions `LALInitBarycenter`, `LALBarycenterEarth` and `LALBarycenter` were also used to compute the time delays between the reference frame of the interferometer and the Solar System barycentre.

Efficient Likelihood Calculation

During early trials of the software, calls to the `LALComputeDetAMResponse()` function were found to account for a large proportion of the computing time when calculating the likelihood. As it was called for every datapoint in the data file on every single proposal of the chain, there was much duplication of the calculation, particularly when the Markov chain remained in one place. To reduce this load, I decided to pre-calculate the values of this function at each datapoint, for a range of values on a grid of resolution 500 points covering the parameter space of ψ. Whenever a value of this function was required within the likelihood calculation, linear interpolation was used to generate a value based on the nearest two points in the lookup table. Since the response function is a smooth function of ψ and the resolution of 500 points is ample to describe it, linear interpolation provides a good estimate of the true value as it would be calculated.

It should also be noted that the analysis requires an estimation of the variance σ_k of each sample B_k in the input data file. In a true analysis of data this is estimated during the heterodyne and downsampling stage described in section 4.1, but here I simply used the variance of the fake noise data which I had selected when creating the datafiles. In order to approximate the performance of the H1 interferometer, a sample variance of 1×10^{-48} was used for these tests, although any variance may be used if the signal strength is varied appropriately to maintain signal to noise ratio. Indeed, within the implementation the dataset is multiplied by a factor of 10^{22} (and the variances by a factor of 10^{44}) to bring the amplitude estimates into a range comparable with that of the $\cos \iota$ parameter. The amplitude estimates are then rescaled to their original range in the post-analysis.

A note on precision

As the study of gravitational waves frequently employs very small numbers, when performing computations it is important to take care that the representation format of these numbers is appropriate for them, and will not introduce errors of approximation which would jeopardise the accuracy of the calculations. As I have mentioned above, quantities of magnitude 10^{-44} and below are encountered here which lie outwith the lower range of the IEEE-854 single precision floating point standard. To prevent the problem of underflow, 64-bit double precision format is used throughout both the MCMC algorithm and the LAL library. The minimum representable normalised non-zero positive number in this system is $2^{-1022} \approx 2.225074 \times 10^{-308}$, which will accommodate our needs at the cost of a slight decrease in speed in comparison to single precision.

In addition to this precaution, when problems might occur related to the dynamic range of the floating point representation, specifically when the h_0 and $\cos \iota$ parameters are reparametrised, a scaling of the h_0 parameter by a factor 10^{22} is performed to bring it into a similar range to $\cos \iota$.

3.1.1 Parameter Estimation

Here, I will use the MCMC code to demonstrate the successful estimation of parameters of an injected signal of sufficient strength. I shall also look at the behaviour of the MCMC code when attempting to detect a signal that is of low signal to noise ratio and discuss these results. Since this is a test of the correct recovery of parameters, I have chosen to start the Markov chain at the very point in parameter space where the signal was injected. This should provide the best possible chance

for the recovery of the signal parameters, with the desired consequence that the behaviour of the algorithm with a low signal to noise ratio in this situation will reveal its fundamental limitations.

To investigate the performance of the algorithm at a range of signal to noise ratio, I have injected signals with varying values of h_0 and used the MCMC code to recover the injected parameters. I shall look at how the probability distribution of these parameters is affected by signal to noise ratio.

Recovery of 6 Parameters From an Injected Pulsar Signal with High SNR

To confirm that the algorithm can indeed recover the parameters of an injected signal, a signal of amplitude $h_0 = 1 \times 10^{-24}$ was injected into a dataset of randomly generated Gaussian noise ($\mu = 0, \sigma = 1 \times 10^{-24}$) with $N = 64\,000$ samples evenly spaced 1 minute apart. The integrated signal to noise ratio in this instance was approximately

$$\mathrm{SNR} \approx \frac{h_0}{\sigma} \times \sqrt{N} = 1 \times \sqrt{64\,000} \approx 252 \tag{3.1}$$

The sampler was asked to generate $1\,000\,000$ samples, which were thinned by a factor of 50 to reduce correlation, yielding $20\,000$ samples in the distribution.

Figure 3.2 shows the marginalised posterior probability density functions as estimated from sample density of the Markov chain. In each parameter, the posterior PDF includes the injected value of the signal parameter, showing that the Markov chain has followed the underlying probability distribution. The chains are shown alongside each histogram to illustrate the random walk nature of the sampling. From inspection of the chains it may be clear that there is some correlation

between different parameters, namely ψ and ϕ_0 and h_0 and $\cos \iota$.

In figure 3.3 the correlation which may be seen in the plots of the Markov chains is made explicit in the 2-D marginalisation of the chain over pairs of parameters. This reveals the correlation which exists between certain pairs of parameters. Particularly evident is the correlation between ψ and ϕ_0, which can be explained by considering the physical meaning of these parameters with respect to the source. The degree of correlation between these parameters is in fact dependent on $\cos \iota$; when $\cos \iota = -1$, the rotation vector $\hat{I}$ is parallel to the line-of-sight vector $\hat{n}$ from the detector to the source, so a rotation of the pulsar about $\hat{I}$, which is a change in ϕ_0, is equivalent to a rotation of $\hat{I}$ with respect to the detector, which is a change of ψ - the two parameters are degenerate in fact. In this case there is no way to distinguish between a change in ϕ_0 and a change in ψ and they are maximally correlated. Likewise, in the case where $\cos \iota = 1$, $\hat{I}$ and $\hat{n}$ are antiparallel, therefore ϕ_0 and ψ are maximally anticorrelated. In the example shown above, $\cos \iota = 0.6$ so the parameters are somewhat anticorrelated. Only in the case $\cos \iota = 0$ is there no correlation between these two parameters. However, this has not posed a problem for the algorithm as the sampling has been effective throughout the mode, with appropriate choice of step sizes.

Recovery of 6 Parameters From an Injected Signal of Low SNR

In this case, the analysis was repeated with the same input data as above, with identical parameters injected apart from the amplitude, which was reduced to $h_0 = 2 \times 10^{-25}$.

At this amplitude, the inference of the parameters is much less precise, and their distributions correspondingly more uncertain. However, the MCMC algorithm
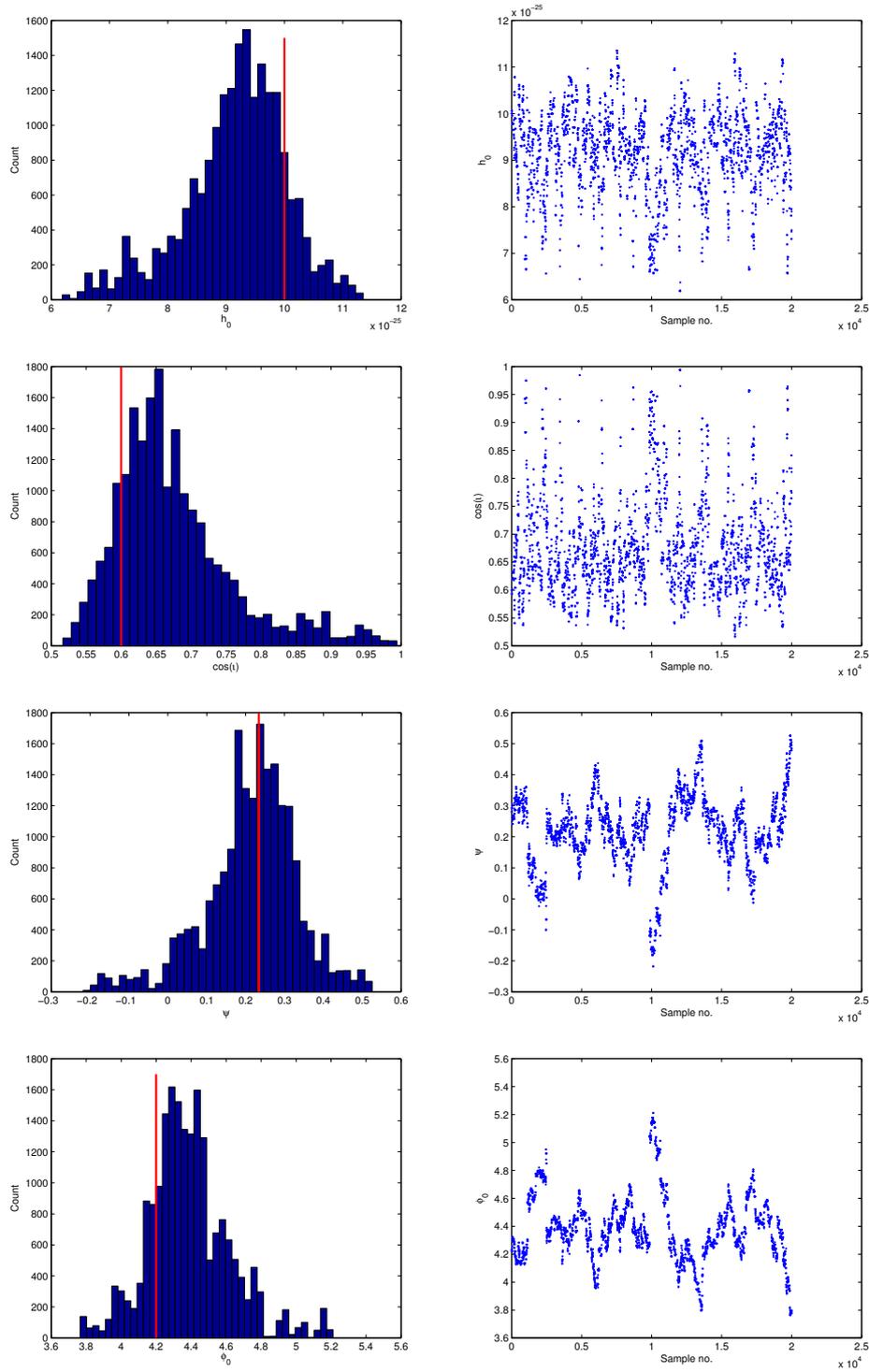
Figure 3.1: [Caption overleaf]
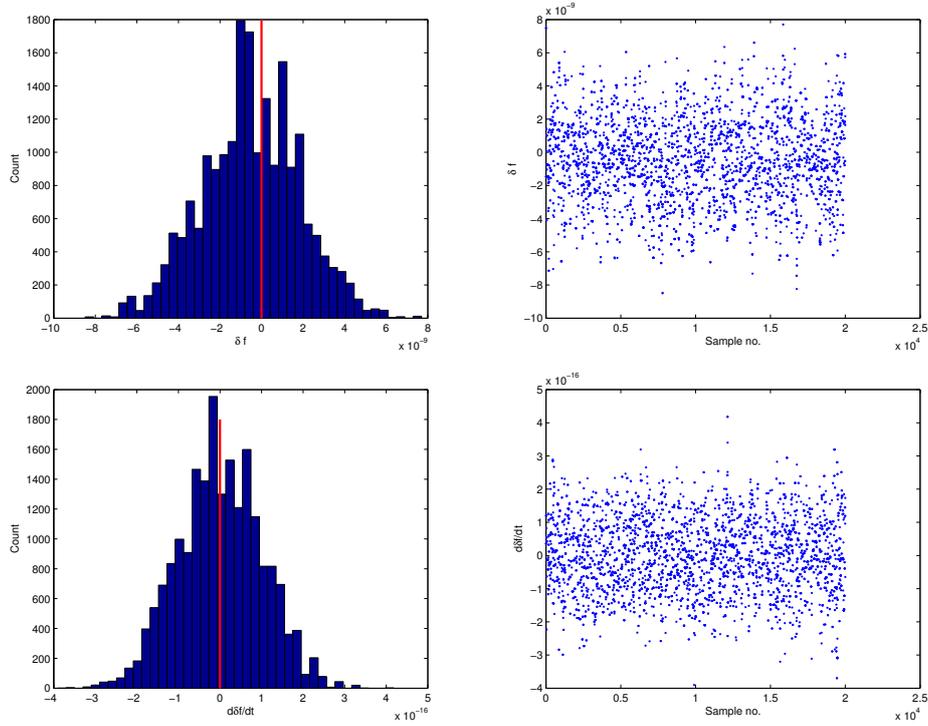
Figure 3.2: Left: Histograms showing marginalised posterior PDF for each of the six parameters (h_0, $\cos\iota$, ψ, ϕ_0, δf and $\dot{\delta f}$). The red line in each histogram represents the injected value of each parameter, the parameter vector for the injection was $\mathbf{a}_{\mathrm{inj}} = (1e-24, 0.6, 0.234, 4.2, 0, 0)$. As expected, each distribution contains the injected value within its main probability mode, assigning a high probability density to it. This shows that the algorithm has successfully sampled from the posterior PDF for this data. Right: The Markov chain which produced the histograms, illustrating the random walk which is taken within parameter space.
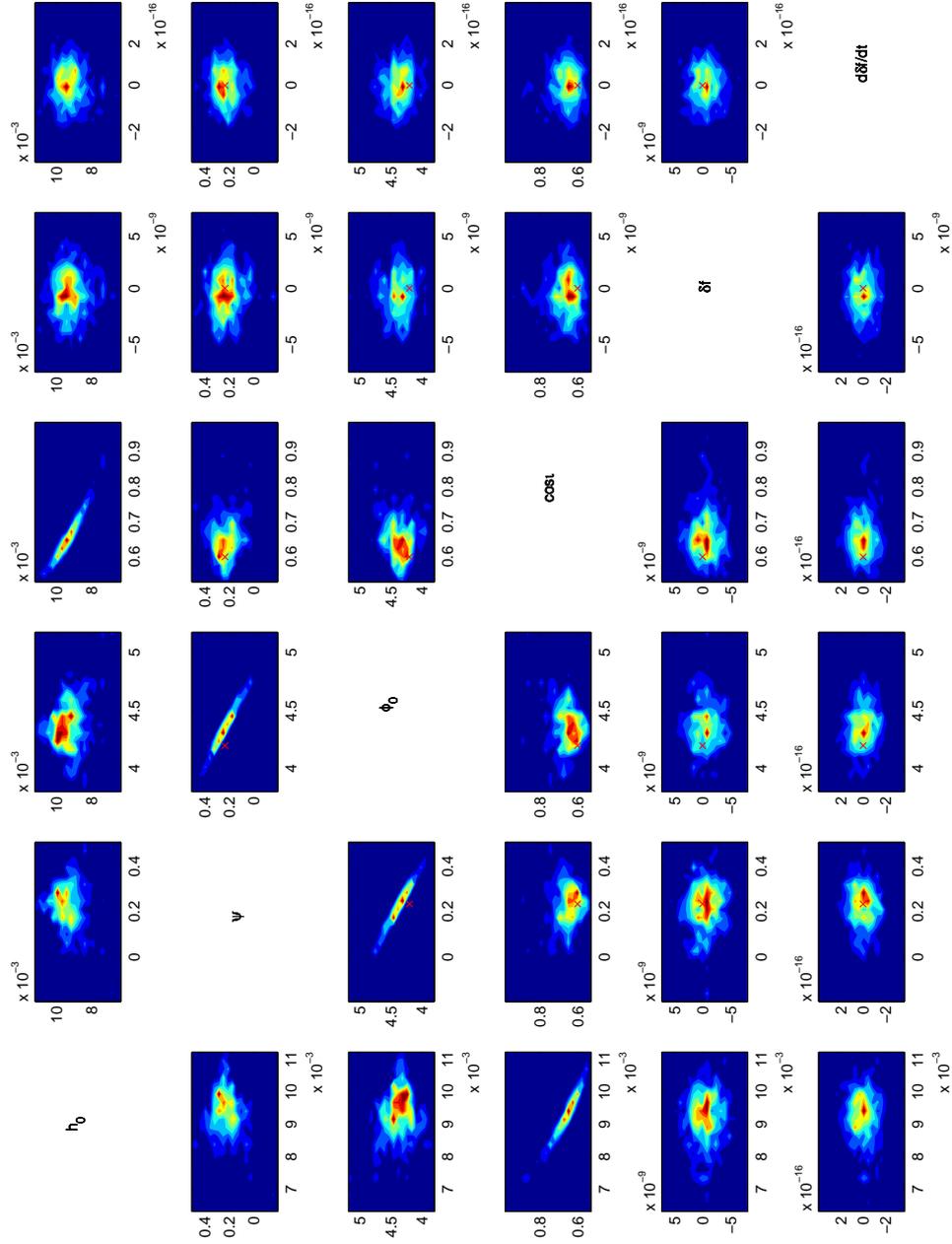
Figure 3.3: The 2-D marginalised PDFs between each pair of parameters, revealing correlations in the h_0,$\cos\iota$ and ψ,ϕ_0 pairs. The injection points are marked with a ×, and lie close to the area of maximum density in every case. The signal was injected into $64\,000$ samples of Gaussian noise with variance $\sigma = 10^{-24}$.
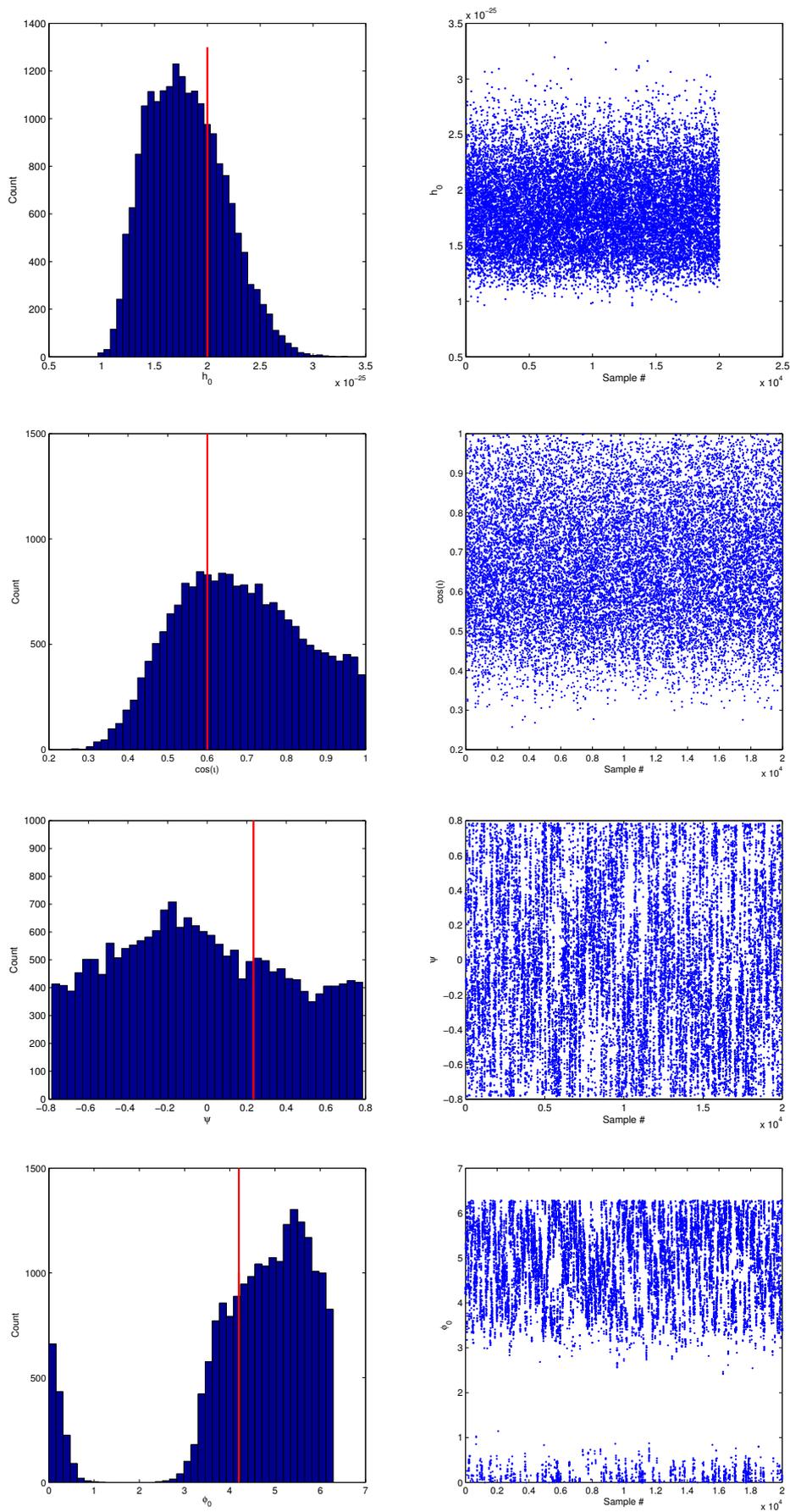
Figure 3.4: [Caption overleaf]
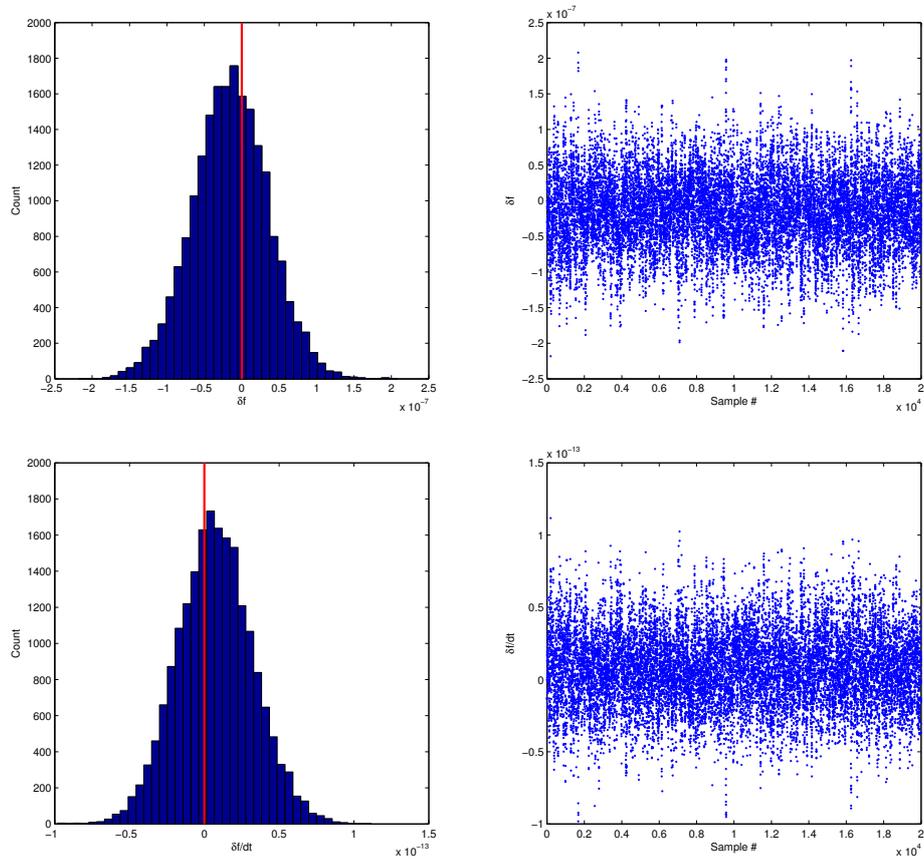
Figure 3.5:   The results of an MCMC sampling of the posterior PDF for an injected signal of amplitude $h_0 = 2 \times 10^{-25}$, all other parameters having been kept equal to those above. The resolution of the parameters is much poorer, with the parameter ψ being barely localised in the marginal PDF. The reason for this is seen in figure 3.6, where the degeneracy with ϕ_0 is apparent. Nevertheless, the amplitude parameter is still recovered quite well with this injection, and the frequency estimate remains correct. In each parameters, the injected value is assigned a high probability density in each case, showing that the distribution based on this noisy data does not contradict the real values of the parameters.

still correctly performs its function of sampling this distribution, which in every case is peaked at or near the injected values of the parameters. In figure 3.5 we see the marginalised parameter estimations from this chain, but figure 3.6 is more revealing. Even though ψ cannot be accurately determined on its own, the marginal distribution of ψ and ϕ_0 remains fairly well confined, with a greater uncertainty in ψ than in ϕ_0.

Estimation of Parameters When the Signal is Undetectable

Recovery of Amplitude

Below a certain limit of SNR, the Markov chain does not remain in the mode of the signal, even when the the mode is the point at which the chain is started. There will always be some small but finite probability of the chain making a jump outside the mode of the signal, and the probability of this being accepted is given by the Metropolis ratio, which can be expressed as

$$p(\mathbf{a}_{\mathrm{sig}} \to \mathbf{a}_n) = \exp\left(\chi^2_{\mathrm{sig}} - \chi^2_n\right) \qquad (3.2)$$

where $\mathbf{a}_{\mathrm{sig}}$ represents the parameters of the signal injected and $\mathbf{a}_n$ is a set of parameters outside the mode.

In white noise of finite length, there is a non-zero probability that there will be a set of parameters which fit the data better than those of the real signal. This is an inevitable consequence of the data following a random distribution when we do not know a priori if there is a signal present. A global maximum of the likelihood will exist in all cases, with some set of parameters which best fit the data, even when there is no signal present. When attempting to fit a signal to this
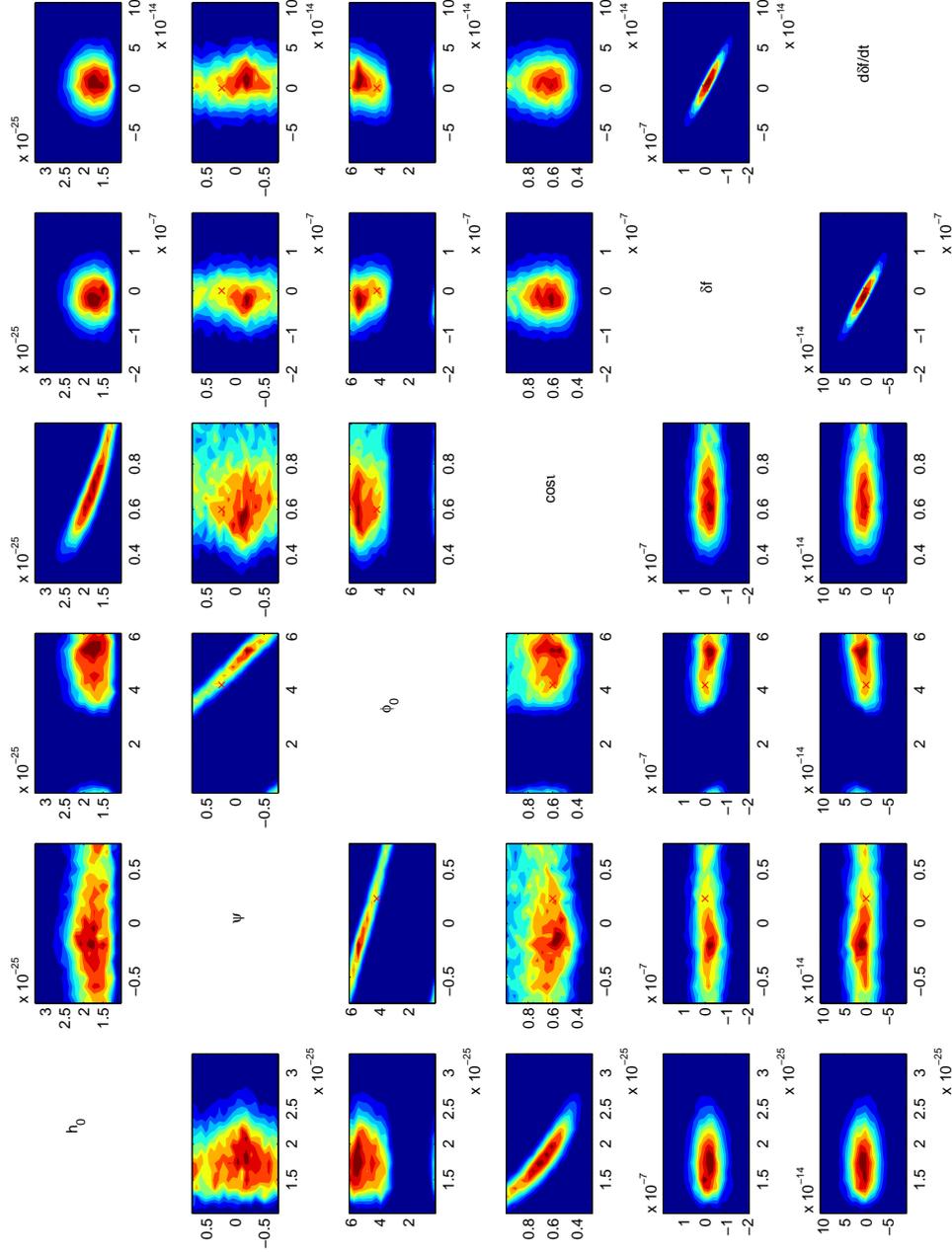
Figure 3.6: As in fig 3.3, the 2-D marginalised PDFs between each pair of parameters for a signal of amplitude $h_0 = 2 \times 10^{-25}$ injected into the same random data as above ($N = 64\,000, \sigma = 1 \times 10^{-24}$). In comparison to the high SNR injection, the parameters are less well-defined and the modes broader. The correlation is more evident between the parameters, and in the non-Gaussian joint marginal PDF of $h_0, \cos \iota$ the curvature of the mode is clearly visible. The injected values of the parameters are marked with a $\times$.

data those parameters or perhaps some other set of locally maximal parameters will be detected with a minimum χ_n^2. If this maximum is not at the point of zero amplitude, then there must be a lower probability at zero than at this point. In the case of finite Gaussian white noise, the distribution of amplitudes fluctuates with changing frequency, just as the time series amplitudes fluctuate, since the Fourier transform of a normal distribution is also a normal distribution. Since there must therefore be a particular frequency with a maximum power, it is evident that at this frequency, the probability distribution with respect to amplitude h_0 must peak at $h_0 > 0$, since a peak at $h_0 = 0$ would be independent of frequency as all points in the model $y_k(t) = 0$.

Now consider if there was a very weak signal injected into this data; so weak that the fit at its parameter vector generates a $\chi_{\text{sig}}^2 > \chi_n^2$. It is clear from the above considerations that for a data set of any non-zero noise level and finite length one can find a signal small enough to fit this criterion. Further more, at amplitudes decreasing below this criterion, it is obvious that an increasing proportion of the parameter space has a χ^2 greater than χ_{sig}^2 since the probability distribution on this space is smooth and continuous. Eventually one reaches an amplitude so low that any proposed jump outside its mode has some reasonable probability of being accepted, and with the probabilistic nature of the MCMC algorithm it is likely that after a number of jumps the chain will have moved away from the signal parameters $\mathbf{a}_{\text{sig}}$. Since the probability of the chain jumping back into such a small area of parameter space is exceedingly low, and the chain cannot be run for an infinite period of time, such a signal may be regarded as undetectable in practice[1].

---

[1]If not in principle - an infinitely long chain would sample the entire search space sufficiently to reveal its mode, although the maximum probability there would likely lie below the global maximum probability.

This is an unavoidable consequence of searching over frequency. Note that in the Bayesian search for known pulsars where frequency is not a parameter, there are a proportion of posterior PDFs which peak away from zero as the target frequency will randomly fall on one of those frequencies at which the noise conspires to produce a non-zero amplitude [26].

When the signal lies below this threshold, then there is no practical way of detecting the signal with the MCMC algorithm, since if the finite chain were to jump into the mode, it would jump out again before long, with a low probability of finding its way back before the duration of the run had finished. If this happens, the marginalised posterior probability distributions recovered from the chain will no longer converge on the injected parameters of the signal. In particular the amplitude will not follow the injections down to $h_{\mathrm{inj}} = 0$ but will reach a lower limit corresponding to the amplitude of noise fluctuations when they are larger than that of the signal.

To examine this tendency towards recovering a greater amplitude than injected I have performed a series of injections and examined the trend of estimated amplitude against that injected. Figure 3.7 shows the results of these trials.

As I have described, there is indeed a minimum amplitude which the algorithm returns, which is independent of injection level below the threshold. This is caused by the chain no longer converging on the signal but instead on a random fluctuation of the noise. Therefore, one cannot set upper limits by marginalising this distribution as there is no way of telling whether a signal is present below the threshold. There is a problem in determining whether the amplitude observed is a result of convergence on a signal or of the excess power in the noise.
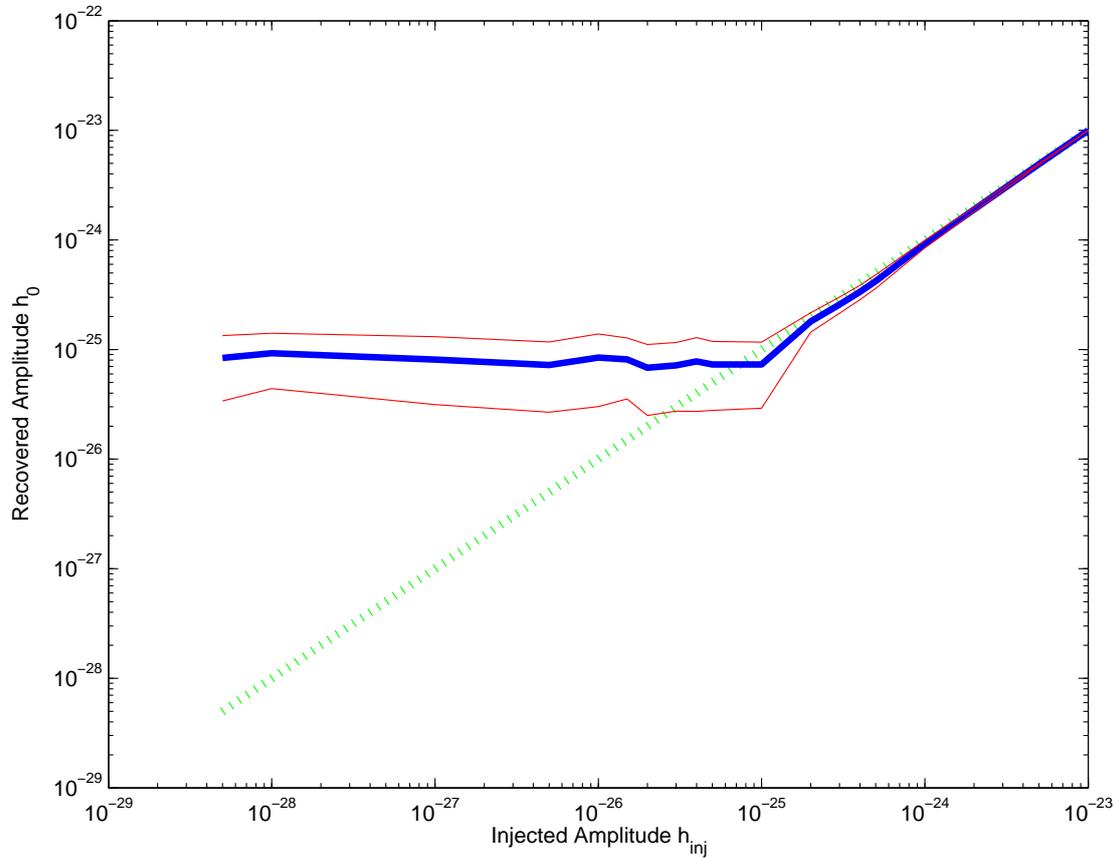
Figure 3.7: This plot shows the results of estimating the value of the h_0 parameter for a range of input values h_{inj}. The solid blue line indicates the mean recovered value, with the $1\,\sigma$ confidence interval above and below marked in red. The dashed green line indicates a 1:1 correspondence between h_{inj} and h_0. At high SNR the amplitude is recovered accurately and the distribution of h_0 is narrow, but as the SNR approaches zero, the estimated distribution of h_0 no longer approximates the injected value but instead converges on a lower limit. This happens when the probability of the signal is no longer significant in comparison to the random fluctuations of the noise, causing the algorithm to favour sampling the random modes of the noise rather than that of the signal.

3.1.2 Determining Convergence

It is a long standing problem in the field of Markov chain Monte Carlo research of how to know when the sampling chain has converged on the target distribution. When the details of the target distribution are not known, then a common means of determining convergence is to run the analysis in many parallel independent chains and examine the resulting sample distributions for a common mean, or other statistical properties [38]. However, since in this case we know that we are looking for a distribution from a pulsar signal, we are able to use that information to test for convergence of the chain. The method I have chosen relies on the nature of the posterior distribution in the frequency parameters, and will now be developed.

Width of frequency modes

Examining the width of frequency modes is a simple way of testing whether we have converged on a pulsar signal or not, as it has been observed that a signal which converges on a pulsar distribution has a very limited frequency distribution. Here we shall observe the nature of the marginal posterior PDF over δf and $\dot{\delta f}$ as the signal to noise ratio varies. In figure 3.8 is shown the variation of the width of the frequency mode from the same dataset which produced figure 3.7.

The plot has a distinctive jump in both modes between $h_0 = 1 \times 10^{-25}$ and $h_0 = 2 \times 10^{-25}$, corresponding to the point in figure 3.7 where the estimated amplitude departs from that injected. This shows where the estimation of the posterior PDF have no longer converged on a signal, and suggests a means to test for convergence automatically. Since in the case of the converged signal,

the frequency mode is contained within one frequency bin of width $1/T_{\text{obs}}$ and the $\dot{\delta f}$ mode within $1/T_{\text{obs}}^{-2}$, an automatic system to check for convergence can be programmed to test for this condition. The linearly decreasing trend of the graph with gradient -1 shows that the width of the frequency mode is inversely proportional to the amplitude.

A theoretical basis for this observation is provided by Bretthorst in [39] and [40], where it is shown that the accuracy to which a monochromatic sinusoid or chirped sinusoid can be estimated is indeed inversely proportional to the amplitude,

$$\delta\omega \approx \frac{\sigma}{nA}\sqrt{\frac{48}{N^3}} \tag{3.3}$$

which is a case very similar to that we have for pulsar signals. Since this will be less than the width of a frequency bin in the Fourier transform of the data, in the case where we have a SNR high enough to detect, we may check for convergence by calculating the width of this distribution in the output chain $2\sigma_{\delta f}$, and testing it to be less than $1/T_{\text{obs}}$. This gives a simple means of checking for convergence on a signal that can be performed automatically by computer as part of the analysis pipeline. In the real search similar criteria were used with the thresholds tuned for use with real data where interference lines can mimic a signal.

This is illustrated by comparing figure 3.9 and 3.10, where the two cases are clearly distinguishable by eye and by a simple machine-applied test of this property.

The lower limits of detectability

It is evident that no search algorithm can detect a signal with vanishingly low SNR, and it is important to characterise the sensitivity of an algorithm when
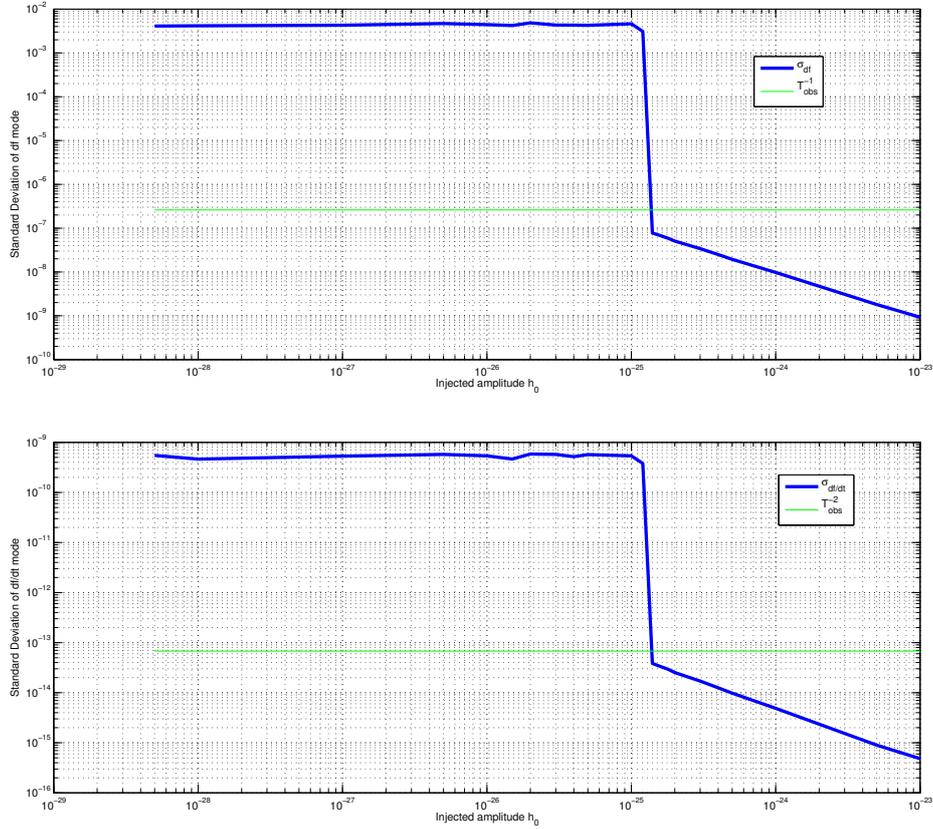
Figure 3.8:   Top and bottom: the standard deviation of the frequency and 1st spindown modes respectively given a range of amplitude injections on a logarithmic schedule. The width of the modes is comparable to the entire parameter space at low SNR, showing the chain has not converged on a signal. At a certain threshold, the situation changes suddenly and the width falls below the threshold indicated by the green line as the reciprocal of the observing time T_{obs}^{-1}, and for the $\dot{\delta f}$ parameter T_{obs}^{-2}, which is the maximum scale of the posterior PDF for a signal. The trend then proceeds proportional to $\frac{1}{h_0}$, as expected from theoretical considerations [39].
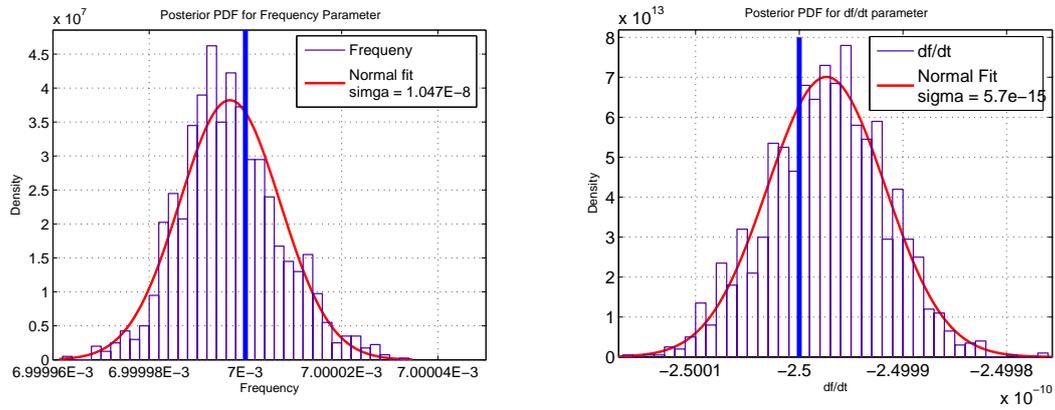
Figure 3.9: The posterior PDFs (bars) for a 2000 sample chain which has converged on the injected values of $\delta f = 7 \times 10^{-3}$ Hz and $\delta \dot{f} = -2.5 \times 10^{-10}$ Hz/s (vertical lines), and the best normal fits, having standard deviations 1.047×10^{-8} Hz and 5.7031×10^{-15} Hz/s respectively.
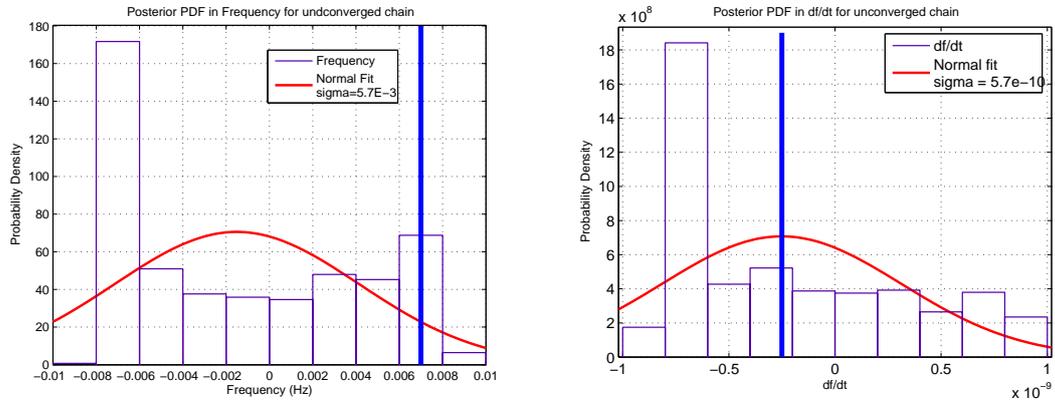
Figure 3.10: The δf and $\delta \dot{f}$ posterior PDFs (bars) for a 2 000 sample chain which failed to converge on the injected parameters (vertical lines). This distribution is clearly distinguishable from the converged case (fig 3.9 using the criteria defined above.

evaluating it for use. We have looked above at the absolute limits of performance of the MCMC search algorithm, where it was started with the chain already in the mode of the signal. This shows us the very lowest SNR that is in principle detectable. However this answers only part of the question of the sensitivity of the MCMC algorithm, as in a realistic situation we do not know in advance where the maximum lies in the parameter space. When performing a *search* we do not have this information, and as a probabilistic algorithm, there is of course a chance that the algorithm will not happen upon them during the course of its finite run.

The sensitivity of the search is determined both by the nature of the signal, and by the length of time that we are prepared to spend running the search algorithm, or rather the number of computational iterations we are willing to devote. The longer the run, the greater probability the chain has of finding the hyper-volume in which the bulk of probability lies in the case of a signal. In the next section I will quantify this probability and show how it can be used to set upper limits on possible signal strength, despite the unreliability of results obtained when no signal is detected.

3.1.3   A Method of Setting Upper Limits

In the situation that an analysis has been performed, and no signal was found in the data, it is useful to be able to set an upper limit on the amplitude of any undetected signal that may be present. This is possible in the Bayesian search algorithm by simply marginalising the poster PDF over all parameters except h_0, then integrating the h_0 distribution upward from $h_0 = 0$ until 95% of the total probability is under the integral. The value of h_0 at this point is the 95% upper

limit on h_0.

$$0.95 = \int_0^{h_{95\%}} p(h_0|\{B_k\}, I)dh_0 \qquad (3.4)$$

In the case of our MCMC algorithm however this procedure cannot be used, since the parameter estimation cannot be relied upon to produce an accurate marginal distribution for the amplitude parameter. There may be a greater total probability lying in an area of low density near the $h_0 = 0$ axis but the Markov chain will tend to fall into one particular non-zero maximum in the density distribution and stay there, since its density there is much greater than that near $h_0 = 0$. Therefore in order to set upper limits on the amplitude of gravitational radiation being present in the data, I have taken a different approach. From the information presented in figures 3.7 and 3.8, we see that there is a certain sharp threshold below which no signal is detectable even with prior information on where to look. At some point in this transition zone, there will be a signal which is detectable in 95% of runs. This is possible since the algorithm is not deterministic, and different random seeds will lead to different Markov Chains. It is possible then to find the point at which we can state with 95% confidence that, having run the algorithm for N iterations, we believe there are no signals of this strength or greater.

Whereas before we looked at the probability of maintaining convergence on a known signal, in this section I will investigate the probability of detecting a signal if we do not know in advance where it lies. This is the more realistic scenario, and knowledge of this probability allows us to set upper limits on gravitational radiation strength in a particular dataset from a particular source. This method was presented at the Gravitational Wave Data Analysis Workshop 9, and was published in its proceedings [41].

Consideration of relevant parameters

If we are to proceed by performing a Monte Carlo simulation of many injected signals and trying to detect them, we need to know what factors are important in determining their detectability. Returning to the likelihood function of equation 2.13,

$$y(t_k; \mathbf{a}) = \left(\frac{1}{4} h_0 F_+(t_k; \psi)(1 + \cos^2 \iota) - \frac{i}{2} h_0 F_\times(t_k; \psi) \cos \iota \right) \exp(i\phi(t_k, \delta f, \dot{\delta f})),$$

$$(3.5)$$

and the accompanying equations 1.18 and 2.14, one can see that the parameters δf, $\dot{\delta f}$ and ϕ_0 are involved merely in the phase evolution of the signal, and do not contribute to the amplitude. Therefore they do not contribute to the detectability of the signal by raising the SNR, since any data run is likely to contain a significant number of cycles, making the initial phase irrelevant.

In 1.18 we see the effect of polarisation on the amplitudes of the $+$ and $\times$ parts of the signal. This can affect the amplitude of the received signal as would be expected. However, this effect varies sinusoidally also, on the timescales at which the detector rotates relative to the fixed source position on the sky. This motion is due to the rotation and orbit of the Earth, so at timescales greater than 1 day the effect of this angle will diminish accordingly as the detector performs full rotations.

So we are left with h_0 and $\cos \iota$ as being parameters which contribute directly to the amplitude and SNR of the signal as observed in a lengthy data set. This means that the injections of the Monte Carlo must be performed at various points on the h_0, $\cos \iota$ plane to examine the variation of detectability at possible relevant points in the parameter space.

The data format for the MCMC code provides the real and imaginary parts of the observations as separate datastreams; the model is calculated for each part, compared and then the two χ^2s summed to provide the total mismatch. This, has the benefit of making the sign of $\cos\iota$ irrelevant in terms of SNR. Since in the real part it is squared anyway, and in the imaginary part a negative $\cos\iota$ and a positive $\cos\iota$ provide the same absolute magnitude to the signal amplitude. Therefore we need only examine the half-range $\cos\iota \in [0,1]$ for setting upper limits, since the lower range $\cos\iota \in [-1,0]$ will be a reflection of this.

Procedure

By repeatedly injecting signals at varying values of h_0 and $\cos\iota$, then running the search to try and detect them, we can build up an empirical detection probability $P(\text{detection}|h_0, \cos\iota, I)$, calculated from the fraction of detections at each point on the plane. By subdividing the plane into a regular grid and repeating the procedure at each point, the variation of this probability is built up. Since we are interested in setting an upper limit on h_0, this must be marginalised numerically over N points on the $\cos\iota$ axis,

$$P(\text{detection}|h_0, I) = \frac{\sum_{\cos\iota} P(\text{detection}|h_0, \cos\iota, I)}{N}.$$

Since this generates a probability function (and not a density) the correct procedure for finding the upper limit here is to simply find the value of h_0 at which this function has the value 0.95, since the probability of there being an undetected

gravitational wave is simply $P(\text{non} - \text{detection}|h_0, I) = 1 - P(\text{detection}|h_0, I)$.

$$P(\text{detection}|h_{95}, I) = 0.95. \tag{3.6}$$

This is effectively using a Monte Carlo approach to evaluate the performance of the algorithm, so this combined approach is known as MCMCMC.

Test with artificial data

To test this procedure, 64 000 samples of white noise were generated with variance $\sigma^2 = 1.0$, uniform sampling rate $\Delta T = 60\,s$. The first timestamp of the data was 751658720 GPS seconds. An artificial signal was used for the injections, with parameters shown in table 3.1.3, and with h_0 and $\cos \iota$ varying.

Parameter	Value
Right Ascension	1.463751648
Declination	-1.208988555
ψ	0.281
ϕ_0	4.234
δf	7×10^{-3} Hz
$\dot{\delta f}$	-2.5×10^{-10}Hz s^{-1}

The upper limit that is found with this method depends upon the number of iterations that are used in the MCMC routine. This is held constant throughout the trials to give an answer for a certain number of iterations. In this case, the number of iterations used was 1 100 000, with 1 000 000 iterations being the length of the burn in, and the following 100 000 used for sampling the distribution. This ratio was arrived at after many trials, where it was found that a longer burn-in

time helped a great deal in searching for a signal, with more of the parameter space being explored.

A series of trial runs was performed at constant $\cos \iota$, which found the rough scale of the transition zone. Then a rectangular grid of 10 points was laid in the $\cos \iota$ direction, and 13 in the h_0 direction. The injections were performed at each of these grid points.

Results of MCMCMC Upper Limits

Figure 3.11 shows the results of the Monte Carlo injections on the $\cos \iota$, h_0 plane. As predicted the results vary strongly with both parameters, with better sensitivity when $|\cos \iota|$ is nearer unity. These results are then marginalised over $\cos \iota$ to produce a distribution on h_0 alone.

In figure 3.12 we see the distribution on h_0 alone after marginalisation. This shows the probability of a signal being detected as a function of amplitude, with all other parameters being averaged out. This allows the upper limit to be stated in terms of h_0 alone. In this case, the false dismissal probability falls to 5% when the injected amplitude reaches $h_0 = 1.44$.

The sensitivity of an algorithm can be characterised by displaying it as a function of the observing time T_{obs} and the noise power spectral density S_h. When this is calculated for the MCMC algorithm, the expression obtained is

$$h_{95\%} = 515.2 \sqrt{\frac{S_h}{T_{\text{obs}}}}, \tag{3.7}$$

with the constant 515.2 being characteristic of the sensitivity of the search.

In section 4.4.1, this figure is placed into context with another set of analyses
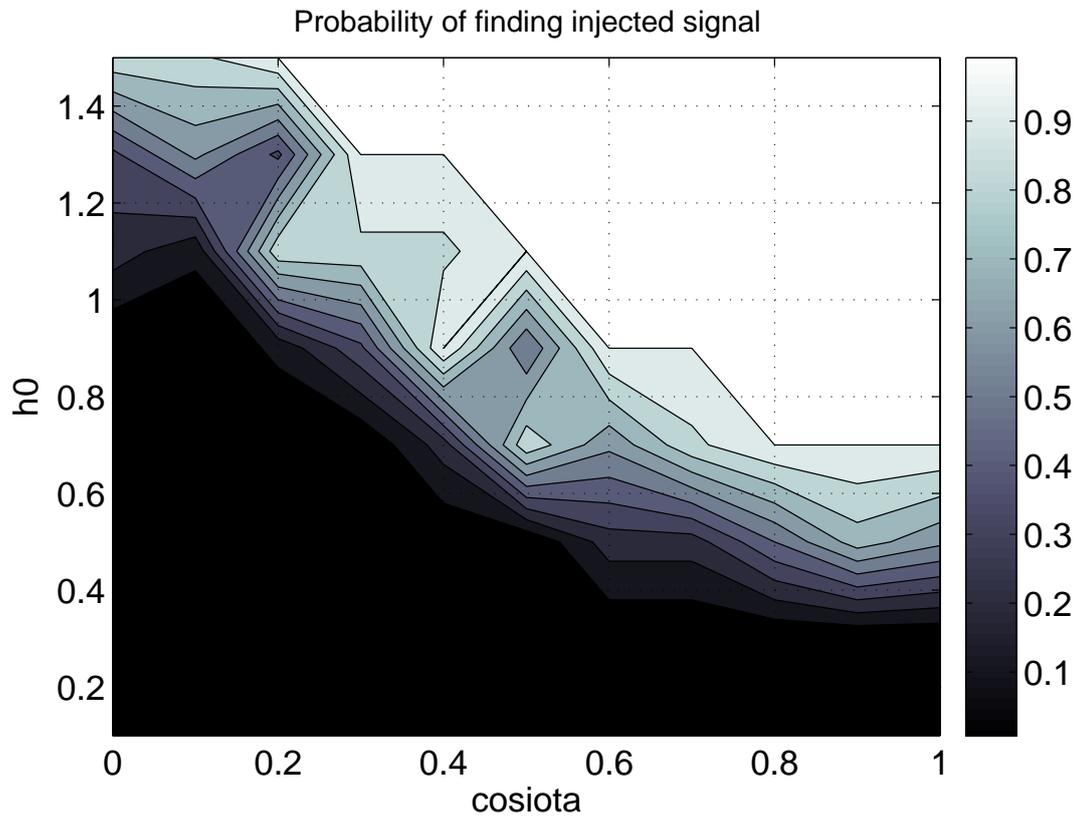
Figure 3.11: The results from the Monte Carlo runs over h_0 and $\cos \iota$, showing that the detection probability does indeed depend on both these parameters. The sudden transition between detectable and undetectable takes place over a narrow range. At larger values of $\cos \iota$, the required amplitude in h_0 is smaller, as additional power is present in the imaginary part of the signal, leading to higher overall SNR.
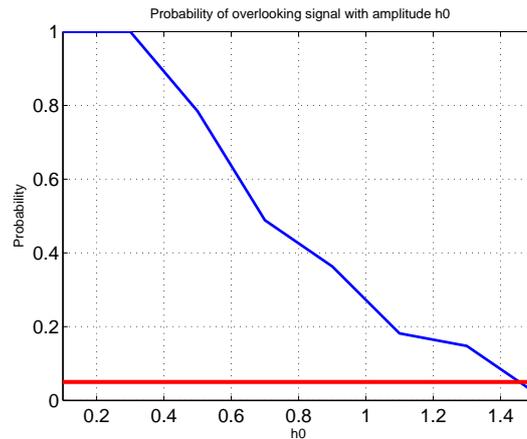
Figure 3.12: Figure 3.11 marginalised over $\cos\iota$ to produce an function of h_0 alone. As can been seen from the diagram, the amplitude at which the detection probability crosses the threshold defined is 1.44.

on real LIGO data, as opposed to the simulations here, and compared with the sensitivity of an exhaustive search. The results presented here are consistent with results seen when using real data, but these indicate that the MCMC algorithm compares poorly in terms of sensitivity with the exhaustive search, being an order of magnitude less sensitive. This will also be discussed further in section 4.4, along with possible regimes in which such a search would be advantageous.

Chapter 4

MCMC Search for Gravitational Waves From a Rotating Neutron Star of Uncertain Frequency

In this chapter the MCMC pulsar algorithm is applied to real data from the LIGO interferometers, with searches performed for the hardware injections in the S3 run, and with a search and upper limit based analysis for a possible neutron star in the remnant of Supernova 1987A. For this analysis a full pipeline was developed to lead from the raw interferometer data to setting upper limits (or detection) and this is documented below.

4.1 Preparation of Interferometer Data

In this section, I will describe the treatment of the LIGO data prior to the analysis taking place. The object of this procedure is to reduce data volume and estimate

the noise in the interferometer across the frequency band of interest, which is necessary for the upper limit procedure.

4.1.1 Heterodyne and Downsampling

The first stage in processing the data is to reduce the data volume, from a sampling frequency of $16\,384$ Hz to $\frac{1}{60}$ Hz, which vastly reduces the number of points that must be calculated in the model, and therefore the speed of the process.

The data from the interferometers are gathered into *frame* files, which are distributed around the world to nodes in the LIGO Datagrid and held at these sites for extended periods while they are used in the searches. These nodes are large computing clusters, which typically have anywhere between 40 and 400 CPUs available for use in the analysis of data in a Beowulf cluster arrangement. This means that the cluster consists of many standalone computers, which are controlled centrally from a head node, in this case with the Condor software [42].

To process the data, a code was developed based on the existing pre-processing code for the time-domain search. However, for the MCMC analysis multiple neighbouring $\frac{1}{60}$ Hz channels of data are required, so there was significant modifications to be made in collecting this data.

To allow the selection of good quality data when the interferometer is in lock, and there is no substantial interference, the extent of a LIGO science run is broken into periods of good data quality and poor data quality and these times are published in a cache file. Using this data, a short script was used to select the appropriate frame files from the collection on disk for the periods of high data quality. One of the advantages of the time-domain analysis is that it can easily

cope with missing segments of data, or non-contiguous sampling periods. This allows data from the entire science run to be used in the analysis, increasing sensitivity over methods which require that a continuous data series is used by virtue of there being more data available.

The locked data segments contain data sampled at ~ 16 kHz, which must be reduced to allow an analysis to take place. The segments are subdivided into one-minute intervals from which each final sample B_k will be drawn. The exact definition of B_k is given below.

For each of the targeted frequencies required (which range over a $4\,\mathrm{Hz}$ window in the search described in section 4.3) the data is heterodyned down to a D.C. signal, by multiplying the data by a sinusoid whose phase evolution matches that of the target, but with opposite sign. The complex heterodyne returns real and imaginary parts of the signal as follows:

$$\mathscr{R}\, y(t_j) = y'(t_j) \times \cos -\Phi(t'_j) \tag{4.1}$$

$$\mathscr{I}\, y(t_j) = y'(t_j) \times \sin -\Phi(t'_j) \tag{4.2}$$

with the phase evolution $\Phi(t_j)$ calculated as a function of frequency f, first and second order spindown parameters $\dot{f}$, $\ddot{f}$ and the position on the sky in equatorial coordinates α, δ. The samples $y(t_j)$ are then equal to those found in equation 2.13, when $y'(t_j)$ represents the original datastream from the interferometer at times t_j. The timestamps at $16\,384$ Hz are modulated according to equation 2.12 where all delays are calculated, as before, using the LAL library barycentring routines.

This heterodyne is equivalent to a translation of the signal in the frequency domain, bringing the instantaneous frequency of the signal down to $0\,\mathrm{Hz}$ throughout

the length of the observations. The components left are a function of the mismatch between the target heterodyne parameters and the true parameters of the signal, and the four parameters h_0, $\cos \iota$, ψ and phi_0.

The data is then filtered to remove unwanted components of the noise from the data. The procedure used here is directly taken from the equivalent data processing used in the time-domain targetted pulsar search documented in [26]. This is achieved with the sequential application of three third-order Butterworth filters (equivalent to a ninth-order filter which is not available in LAL), which have a flat frequency response in the pass-band, and a corner frequency of $0.5\,\text{Hz}$, which was chosen as an acceptably low pass-band without causing ringing of the filter, which occurs with tighter filters of this nature [43].

With the possible signal now centred at $0\,\text{Hz}$, the information needed to analyse the $\pm\frac{1}{120}\,\text{Hz}$ around it can, by Nyquist's Theorem, be encapsulated in a data series with sampling interval $\Delta T = \frac{1}{60}\,\text{Hz}$. With this in mind, the data is downsampled to that frequency by simply averaging each 1 minute stretch of $16\,\text{kHz}$ heterodyned data to produce a point estimate for that time period.

$$B_k = B(t_k) = \frac{1}{60 \times 16\,384} \sum_{j=16\,384k}^{16\,384(k+1)} y(t_j) \tag{4.3}$$

This procedure is repeated for each of the $\frac{1}{60}\,\text{Hz}$ bands into which the frequency window is divided. An overlap of 50% is used between each of these bands to ensure that no power is lost to the filter at the edge of each frequency band, so the central heterodyne frequencies of these bands are spaced $\frac{1}{120}\,\text{Hz}$ apart. The real and imaginary parts of the result are stored separately to be recorded with the noise estimation in the output file. Note that the use of both the real and

imaginary parts of the sample implies that frequencies up to $\frac{1}{60}$ Hz are accurately represented, and the Nyquist limit present with real-only samples does not apply. In effect, we have twice the information content, and so can represent frequencies twice as large.

Noise Estimation

In addition to the point estimate of the complex amplitudes, the variance of this estimate at each time stamp is also required for the likelihood function

$$p(B_k|\mathbf{a}, \sigma_k) \propto \exp\left[-\sum_k \frac{1}{2\sigma_k} \left(y(\mathbf{a}; t_k) - B_k \right)^2 \right]. \tag{4.4}$$

This noise estimate σ_k is produced by calculating the noise power in the 4 Hz window from which the $\frac{1}{60}$ Hz data streams are taken. By this method the noise estimation is the same in each of the $\frac{1}{60}$ Hz bands, which is a suitable approximation when the noise power spectrum is flat in the region in question.

The filtering procedure used here is identical to that described immediately above, with the exception that the Butterworth filters have a corner frequency of 4 Hz, instead of $\frac{1}{60}$ Hz. This preserves the noise in only the 4 Hz band of interest. The 60 seconds of data this time is subdivided into M segments of length 0.25 seconds. Within each segment the data is heterodyned and filtered, and the mean real and imaginary parts of the samples in this segment are calculated. This yields M averaged samples μ_j of the amplitude for each datapoint B_k. As the program works its way through the chunks, it also computes a mean value for the entire 60

seconds of data μ_k. The variance is then calculated as

$$\sigma_k^2 = \frac{1}{M(M-1)} \sum_j^M (\mu_k - \mu_j)^2 \qquad (4.5)$$

which gives the variance of each B_k from the noise power spectral density in its neighbouring window.

This quantity is used in the computation of upper limits, where having the same σ_k for each of the B_k in all the frequency bands allows the results of the Monte Carlo injections to be applicable to the whole 4 Hz band, provided the noise spectrum in that 4 Hz is flat.

4.2   Test of the algorithm on hardware injections into the LIGO Hanford interferometer during S3

Until this point the algorithm has been tested on simulated data which has the ideal characteristics of a Gaussian distribution, stationarity and whiteness. With the provision of injected signals into the LIGO interferometers during the S3 science run, we had the opportunity to test the algorithm to determine if it could detect the signals in the presence of realistic LIGO noise.

The LIGO S3 run

The third science run (S3) of LIGO took place between the 31 October 2003 and 9 January 2004, a total observation time of 71 days. For the duration of the science
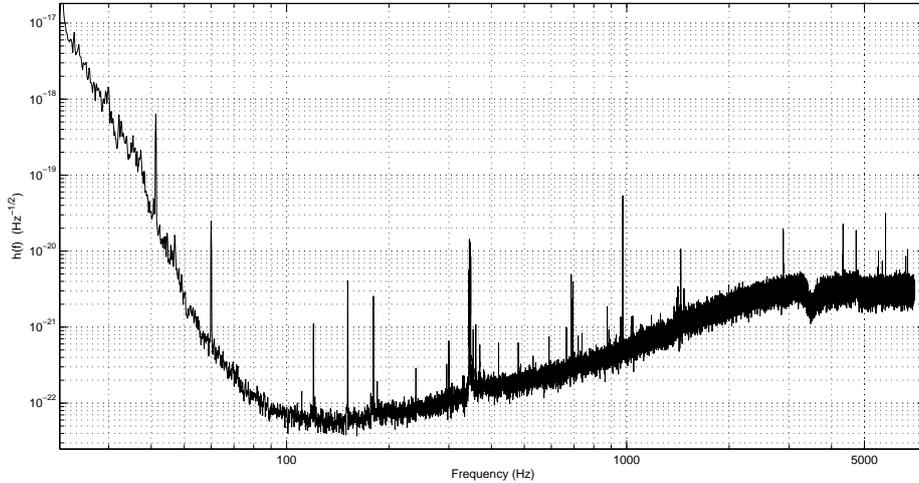
Figure 4.1: Plot showing the strain sensitivity during the S3 science run of the 4 km interferometer at LIGO Hanford Observatory as a function of frequency.

run there were 10 artificial pulsar signals, named PSR0 to PSR9, injected into the data at varying signal to noise ratios to allow verification of analysis codes on interferometer data. From this time period, 25 days of data from the Hanford 4km observatory were selected for the analysis by taking the locked data segments of good data quality only. This did not present a continuous stretch of data, but as the likelihood is calculated in the time domain this did not pose a problem. The data that was used after heterodyning and calibration for injected pulsar PSR0 is presented in figure 4.2.

In this case the data was downsampled to a rate of one complex sample per 30 minutes, instead of 60 seconds, as the frequency and spindown parameters of the injections were known exactly and were transformed so that the signal would lie at $\delta f = 0$, $\dot{\delta f} = 0$ after heterodyne. The estimation of the noise was corrected accordingly to be calculated over the longer period. This lowers the data volume
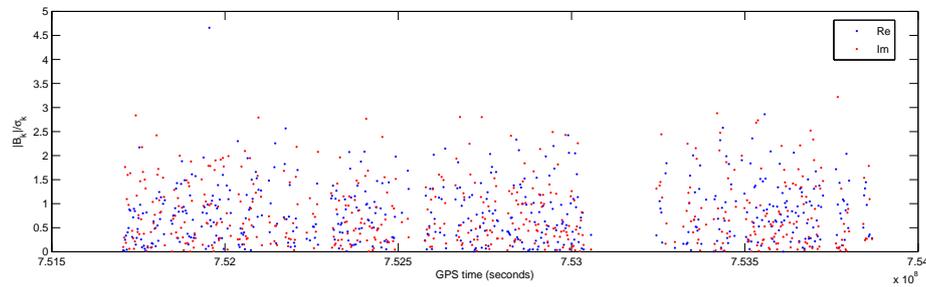
Figure 4.2: A plot of the data used in calculating the likelihood for PSR0 against its observation timestamps t_k. Real and imaginary parts are shown in blue and red respectively as $|B_k|/\sigma_k$. This data closely approximates a normal distribution, as shown in fig 4.3.
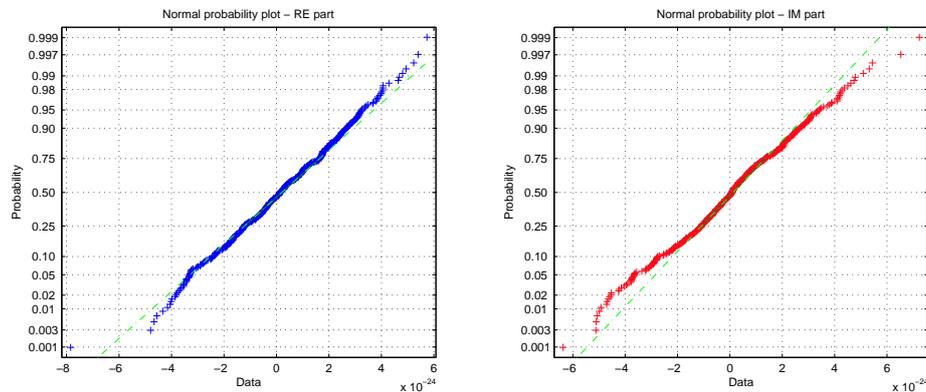
Figure 4.3: A normal probability profile plot of the amplitude of points in the S3 PSR 0 data, showing their distribution of probability compared to a true normal distribution. The green diagonal line indicates a perfect fit to the distribution; the blue and red points again represent the real and imaginary parts of the signal. The bulk of the data points lie on or close to the line, with some outliers at the tails of the distribution where the probability is low.

Name	Rt. Ascension α	Declination δ	Frequency (Hz)	$\frac{df}{dt}$ (Hz/s)
PSR0	1.2488167344	-0.9811802248631	265.577105204882	-4.15E-12
PSR1	0.6526458320410	-0.514042405674	849.083296185424	-3.00E-10
PSR2	3.756928840159	0.0601089582678	575.163573003883	-1.37E-13
PSR3	3.113188712215	-0.5835788033987	108.857159430109	-1.46E-17
PSR4	4.886706853676	-0.2175836463767	1403.16333096977	-2.54E-08
PSR5	5.281831296225	-1.463269033207	52.8083243572857	-4.03E-18
PSR6	6.261385268932	-1.141840210277	148.719025725157	-6.73E-09
PSR7	3.899512715971	-0.3569308339044	1220.97958108098	-1.12E-09
PSR8	6.132905165784	-0.5832631506298	194.308318509895	-8.65E-09
PSR9	3.471208242864	1.32103253788	763.84731653774	-1.45E-17

Table 4.1: The position and frequency information for the ten injected signals in LIGO S3 data.

and increases the speed of the MCMC routine, which must calculate the likelihood function for every data point, so runtime scales as $O(N)$ where N is the number of samples. This data rate was decided upon from the time-domain search, which used it for analysis of these injections also, based on the stationarity of the noise over that time period.

The runs themselves were conducted using a total of 1 000 000 iterations and a burn-in period of 160 000 iterations. The parameters of the 10 injections which were searched over are shown in table 4.1. The epoch of all the signals was GPS second 751680013.0. Frequency derivatives of order 2 and above were all zero.

Pulsars 0-4 were "public", with their parameters being known in advance whereas pulsars 5-9 were "private" and their parameters kept a secret until the end of the science run. In every case the MCMC chain was started at a random point in parameter space, so knowing the true values of PSR0-PSR4 did not assist the estimation of these parameters anyway.

4.2.1 Results from the S3 Injections

In table 4.2 the results of the MCMC runs are shown numerically. I have listed the mean value of the PDF in each case, with a 1σ error estimate on this value if the distribution was approximately normal.

It is clear that that MCMC algorithm performs better when the amplitude of the signal is greatest. I would judge that in injections 3, 4, 8 and 9 the parameters were recovered fully, despite small offsets in frequency and spindown, and these indeed are the loudest injections. In pulsars 1, 2, 5, 6 and 7, the signal was located by the search, as evidenced by the convergence of the frequency parameters, but there are errors in the estimation of the other parameters, whose uncertainties grow larger as SNR falls. In examination of the 6 marginal PDFs of PSR0 (figure 4.4) some evidence of the signal may be seen by the human eye, but it is clearly not strong, with $\cos\iota$, ψ and ϕ_0 filling their respective prior ranges. The criteria outlined above in section 3.1.2 for the quantifiable detection of convergence are not met with this result, that is $T_{\mathrm{obs}}^{-1} = 4.62 \times 10^{-7} < 2\sigma_f$ and $T_{\mathrm{obs}}^{-2} = 2.14 \times 10^{-13} < 2\sigma_{\dot{f}}$. Therefore, these signals were not detected.

This demonstrates once more the utility of MCMC for the recovery and estimation of parameters if the signal is strong enough to attract the chain as it samples the posterior space. As expected, the results become less reliable at lower signal strengths, but the test of convergence based on frequency mode width can identify when there is not a reliable detection.

PSR Name	h_0	ψ (rads)	ϕ_0 (rads)	$\cos\iota$	δf (Hz)	$\dot{\delta f}$ (Hz s^{-1})
PSR0	1.39×10^{-24}	0.553	3.78	0.01	4.0×10^{-8}	6.87×10^{-11}
PSR0 inj.	$\mathbf{9.38 \times 10^{-25}}$	**0.652**	**2.66**	**0.795**	**0.0**	**0.0**
PSR1	$(1.40 \pm 0.3) \times 10^{-23}$	-1.377 ± 0.16	3.76 ± 0.47	0.27 ± 0.17	$(2.6 \pm 1.45) \times 10^{-7}$	$(-0.95 \pm 2.0) \times 10^{-12}$
PSR1 inj.	$\mathbf{8.49 \times 10^{-24}}$	**0.35**	**1.28**	**0.464**	**0.0**	**0.0**
PSR2	$(1.12 \pm 0.28) \times 10^{-23}$	-0.123 ± 0.28	4.3 ± 0.57	-0.598 ± 0.15	$(2.57 \pm 9.49) \times 10^{-9}$	$(-2.6 \pm 2.9) \times 10^{-25}$
PSR2 inj.	$\mathbf{1.562 \times 10^{-23}}$	**-0.222**	**4.034**	**-0.93**	**0.0**	**0.0**
PSR3	$(6.16 \pm 0.1) \times 10^{-23}$	0.4023 ± 0.008	5.53 ± 0.05	-0.0750 ± 0.008	$(-0.2643 \pm 1.4) \times 10^{-8}$	$(0.878 \pm 1.00) \times 10^{-14}$
PSR3 inj.	$\mathbf{6.16 \times 10^{-23}}$	**0.444**	**5.53**	**-0.081**	**0.0**	**0.0**
PSR4	$(9.872 \pm 0.053) \times 10^{-22}$	-0.721 ± 0.004	4.825 ± 0.012	0.295 ± 0.004	$(-3.063 \pm 3.58) \times 10^{-9}$	$(2.06 \pm 3.26) \times 10^{-15}$
PSR4 inj.	$\mathbf{1.00 \times 10^{-21}}$	**-0.648**	**4.83**	**0.2773**	**0.0**	**0.0**
PSR5	2.26×10^{-23}	0.066	2.25	0.185	-3.06×10^{-7}	1.52×10^{-10}
PSR5 inj.	$\mathbf{1.83 \times 10^{-23}}$	**-0.364**	**2.23**	**0.463**	**0**	**0**
PSR6	$(4.941 \pm 0.67) \times 10^{-24}$	0.3147 ± 0.07	0.403	-0.179 ± 0.076	$(7.284 \pm 10) \times 10^{-8}$	$(-8.06 \pm 9.26) \times 10^{-14}$
PSR6 inj.	$\mathbf{5.245 \times 10^{-24}}$	**0.471**	**0.972**	**-0.154**	**0**	**0**
PSR7	3.728×10^{-22}	0.742 ± 0.413	4.742 ± 0.544	0.430	$(-5.289 \pm 6.37) \times 10^{-8}$	$(-6.887 \pm 5.81) \times 10^{-14}$
PSR7 inj.	$\mathbf{2.80 \times 10^{-22}}$	**0.512**	**5.245**	**0.757**	**0**	**0**
PSR8	$(5.985 \pm 0.078) \times 10^{-23}$	-1.470 ± 0.006	2.646 ± 0.025	0.0712 ± 0.006	9.125×10^{-9}	-6.58×10^{-15}
PSR8 inj.	$\mathbf{6.022 \times 10^{-23}}$	**0.170**	**5.890**	**0.0740**	**0**	**0**
PSR9	$(1.615 \pm 0.029) \times 10^{-22}$	0.0979 ± 0.024	1.048 ± 0.050	-0.619 ± 0.014	$(-4.75 \pm 0.5) \times 10^{-8}$	$(8.889 \pm 4.44) \times 10^{-14}$
PSR9 inj.	$\mathbf{1.612 \times 10^{-22}}$	**-0.0086**	**1.0078**	**-0.619**	**0**	**0**

Table 4.2: The results of running the MCMC search on the injected pulsars in the LIGO Hanford 4km data from S3. Injected values are shown in **bold** with results in plain text above. Where the estimated PDF was approximately Gaussian a 1σ error margin is shown
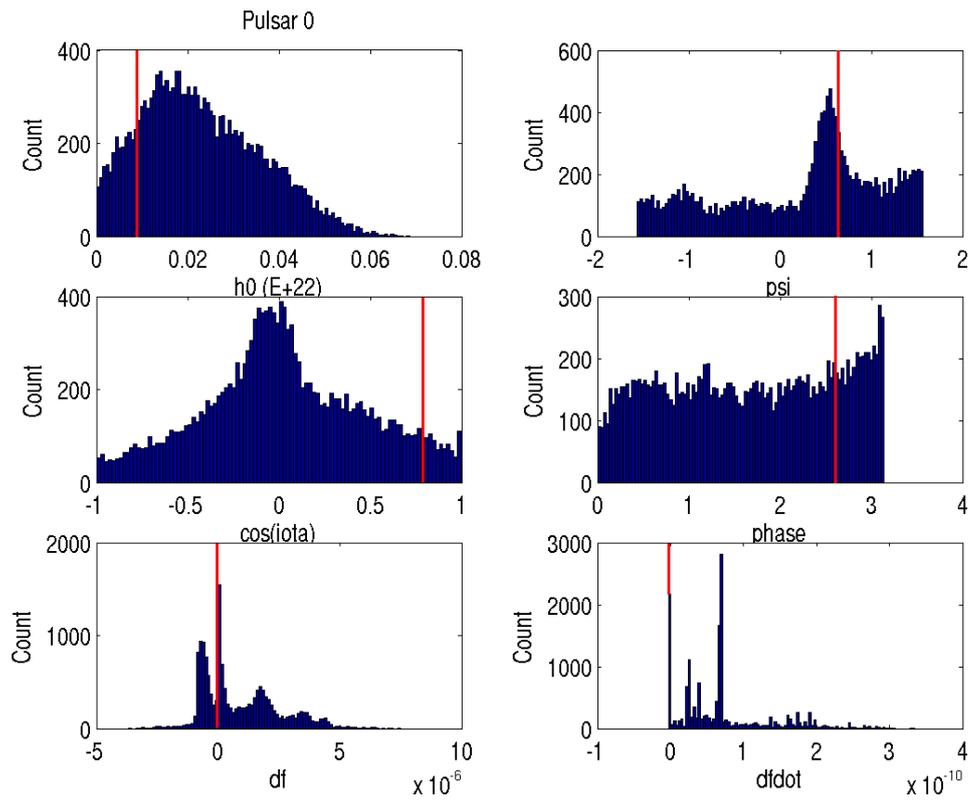
Figure 4.4: Marginal posterior PDFs for the six parameters of PSR0, injected into the Hanford interferometer during S3.

4.3 A Search for a Remnant from Supernova 1987A Using LIGO S3 Data

I will now proceed to describe an astronomical search for a candidate source in the remnant of SN1987a, and why it is amenable to the application of a MCMC search routine.

4.3.1 Review of Supernova 1987A

On February 22nd 1987, the brightest supernova to be observed in 383 years appeared in the Large Magellanic cloud approximately $168\,000$ light years distant. The progenitor star is thought to have been the blue supergiant Sk $-69°202a$, with a stellar mass of between $16M_\odot$ and $22M_\odot$. This would make it highly likely to produce a neutron star upon core collapse, rather than a white dwarf or black hole,

Figure 4.5: Remnant of SN1987A from Hubble Telescope.
Image credit: NASA, ESA, P. Challis and R. Kirshner

and this makes it a possibility that a pulsar was produced [44].

Searches in the electromagnetic spectrum conducted for such an object in the remaining nebula have proven inconclusive, although some evidence did arise for an optical pulsar with an emission modulation at a frequency of 467.5 Hz from the use of high speed photometry [35, 45]. Since some other known optical pulsars are also the result of relatively recent supernovae, such as the Crab and Vela pulsars,

there are no a priori reasons to rule out the plausibility of the claim. The possible signal was intermittently observed between 1992 and 1996 with varying visibility in the data. Since 1996 however the candidate pulsar seems to have disappeared altogether, and while this could be explained by the presence of the nebula and opaque material in the neighbourhood of the remnant, the evidence is not sufficient to conclude for definite the existence of a pulsar.

Nevertheless, the possible existence of such a young pulsar which may possess a large quadrupole moment makes it an interesting target for gravitational wave searches. If we were to assume that the 467.5 Hz modulations were produced by the rotation of the neutron star, as seems likely, then triaxial gravitational wave emission would occur at 935 Hz. The pulsar had a reported spindown rate of $-(2 - 3) \times 10^{-10}\,\mathrm{Hz\,s^{-1}}$ during the period of observation, however this may have changed in the years since and so the current rotational frequency may not be precisely known. For this reason the object is an attractive target for a search which is targeted on a known accurate region of the sky but has a range of uncertainty in which the frequency might lie. Such a target might be addressed by the MCMC algorithm which has been designed to operate within these very bounds.

The searched parmeters were, therefore, the usual h_0, ψ, ϕ_0, $\cos \iota$, δf and $\dot{\delta f}$. The astrophysical parameters of the source which were used in the heterodyning of the data are shown in table 4.3. Since the current spindown and frequency are unknown, a central frequency was chosen at the original observed modulation frequency of the source (doubled to give gravitational wave frequency) and zero spindown was used in the heterodyne stage.

Additionally, the LIGO spectrum at the frequency of interest is very nearly flat, with no bias toward either end of the frequency range (see figure 4.6), and has

RA	05:35:28.03
DEC	-69:16:11.79
f_s	935 Hz
$\dot{f}_s$	0.0 Hzs^{-1}
Epoch T_0	751680013.0

Table 4.3: The position and frequency information used in the heterodyning of data for the SN1987a search.

no obvious interference lines from injections or harmonics of the mains electricity frequency present in the region, so the process of noise estimation over a 4 Hz window is expected to produce a suitable value of the noise.

For this search I have once again used data from the LIGO Hanford interferometer, as it had a higher duty cycle than the Livingston interferometer during the S3 run. Unlike the injections, however, the data used spanned a period of 69 days, with a total of 63 960 samples with sampling period set at 1 minute, interspersed with periods of no samples where the interferometer was out of lock or had data quality warning flags set. There were some greatly outlying samples produced by the ringing of the filters used when the amplitude changed suddenly, and the greatest of these were removed. There were, however, some remaining outlying data points which did not fit a normal probability distribution, as shown in figure 4.8, as a typical example. It is difficult to justify removing points such as these from the data in order to make it follow a more Gaussian distribution - any hard cutoff imposed will tend to skew the distribution in the opposite direction and produce uncharacteristically small tails. Selecting points to discard at random could produce a Gaussian distribution but involves throwing away some data but keeping others with no criterion for choosing between them. It was decided to keep the remaining outliers and operate the search as normal. The data in the
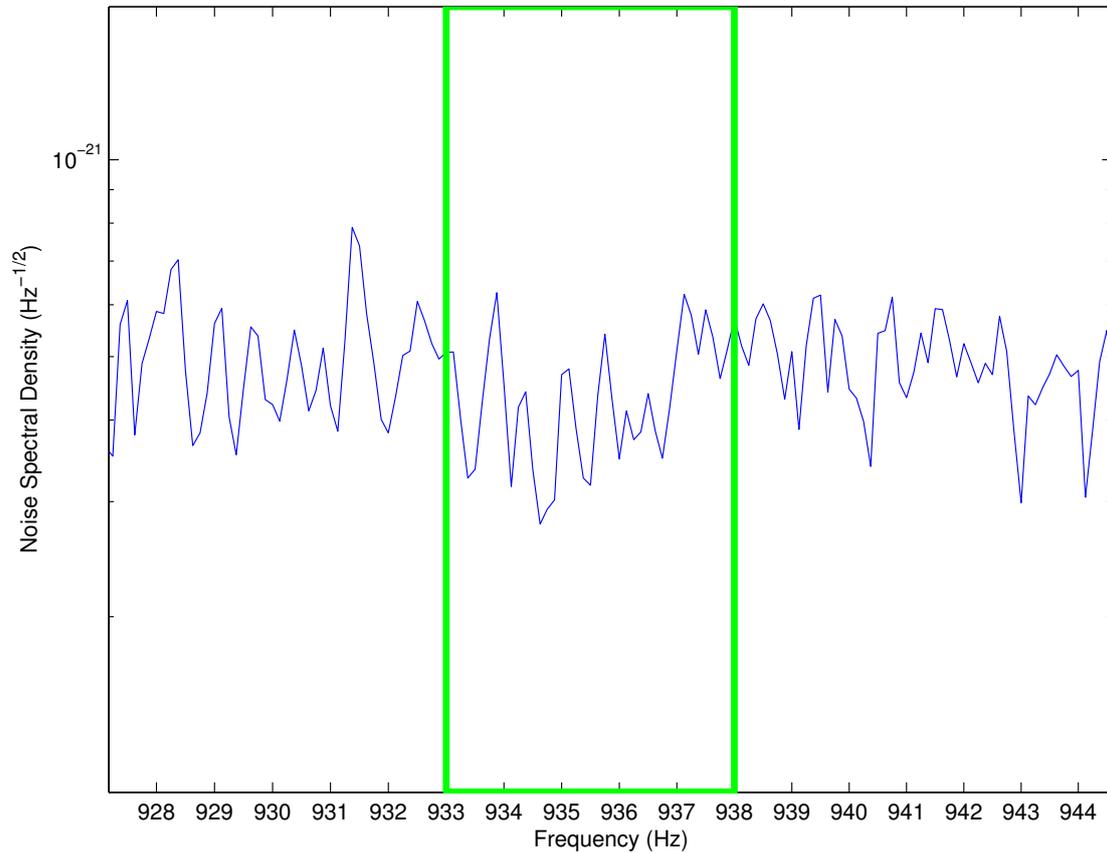
Figure 4.6: The published spectrum from LIGO Hanford 4km Observatory of strain sensitivity averaged over the S3 run, showing the frequency window which is to be searched. There are no strong lines present in this region of the spectrum, so the approximation that the band is flat is appropriate across its width, with no systematic bias toward either end.
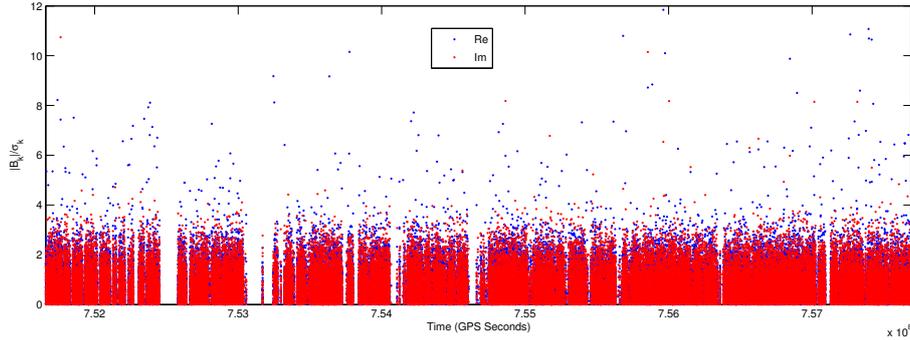
Figure 4.7:   The data used in calculating upper limits for SN1987a, taken from the LIGO H1 interferometer during the S3 run.

first search window, including the outliers, is depicted in figure 4.7 as a typical example of that used in the search. This data was then processed using the full search and upper limit pipeline defined below.

4.3.2   Pipeline for Searching and Setting Upper Limits

The 4 Hz window over which the search was performed was divided into 479 frequency channels, each with a width of $\frac{1}{60}$ Hz, with an overlap of 50% between adjacent channels and the central channel placed exactly on 935 Hz, so as to fully cover the band. Each of these channels could then be searched independently in parallel by using the LIGO Datagrid cluster at Caltech. Each search was performed with a burn-in of length 1 000 000 iterations followed by a sampling period of 100 000 iterations, as was the case in the artificial search above. This was true also for the upper limit finding stage of the pipeline where injections were made in the data.

In the search for an astrophysical pulsar signal in the current generation of interferometers, it is unlikely that a direct detection will be made. So when applying
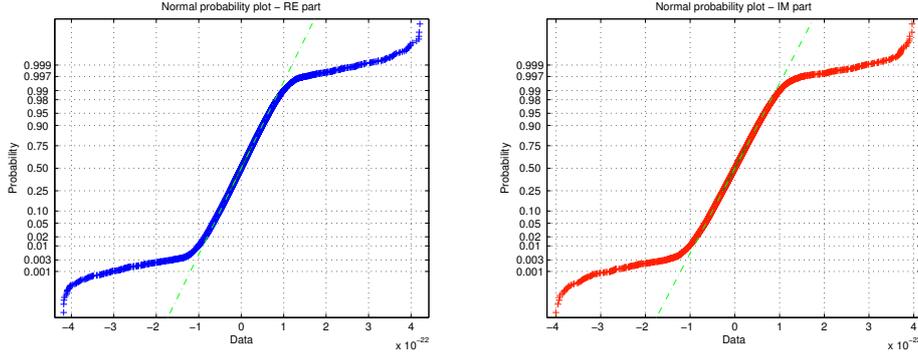
Figure 4.8: Normal probability plot showing the outlying datapoints in the distribution of B_k for the SN1987A data band. Real and imaginary parts are shown in blue and red respectively.

the MCMC search routine to a possible source in SN1987A I have used the upper limit estimation procedure outlined in 3.1.3 in combination with a search similar to that carried out on the injections to produce the results. This brings the steps outlined above into a single combined pipeline based on the MCMC search, which is represented schematically as a flow diagram in figure 4.9.

When the pipeline was fully processed, the posterior PDFs from each individual frequency band was post-processed by a small script which checked the convergence criteria against the output, so as to determine if a possible signal had been observed.

Based on the results of trial runs of injected signals, the criteria used for detecting convergence of the chain were again based on the frequency posterior width, which must meet the criteria $2\sigma_{\delta f} < 5 \times 10^{-8}$ Hz and $2\sigma_{\dot{\delta} f} < 10^{-14}$ Hzs^{-1}.
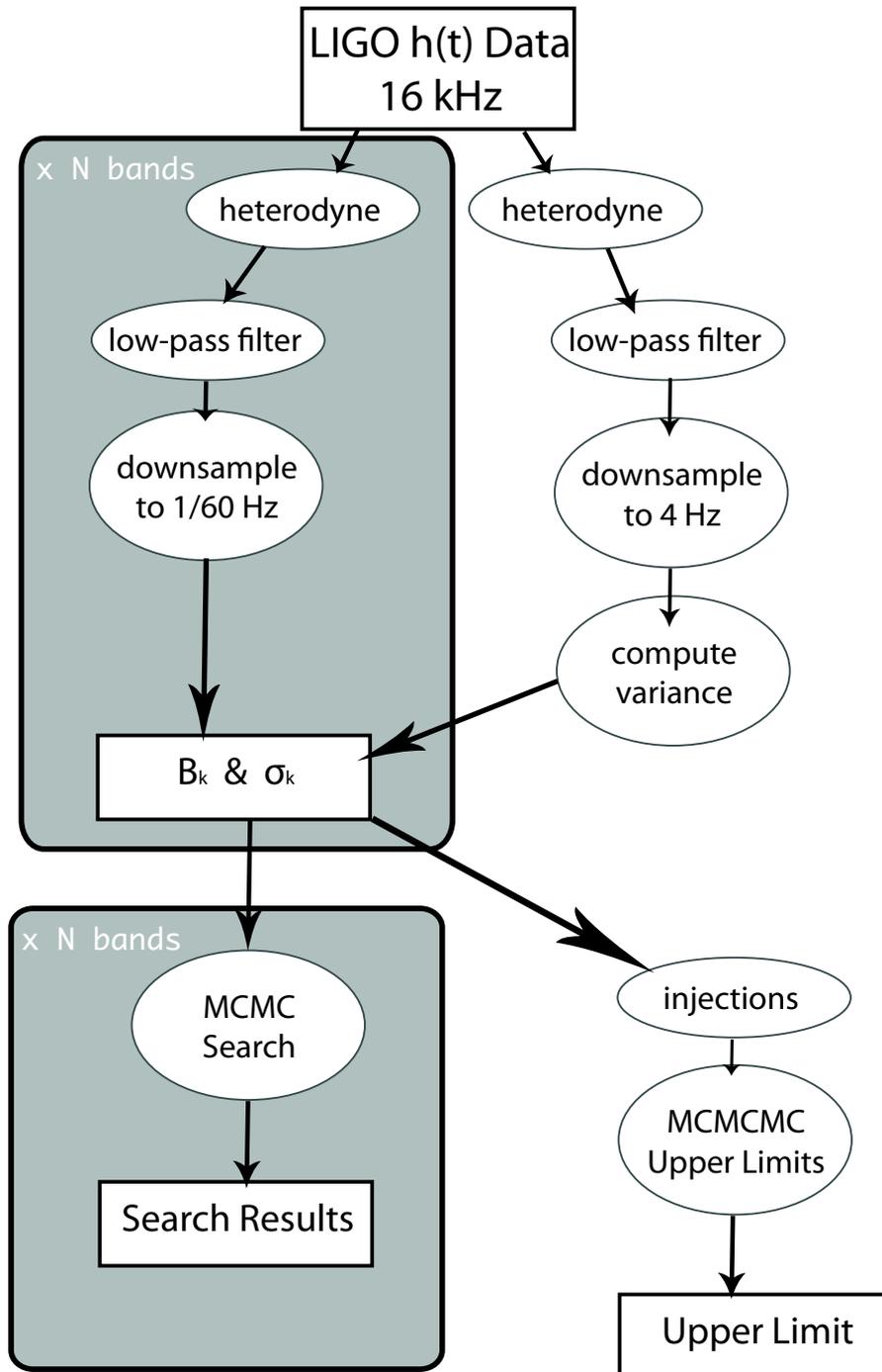
Figure 4.9:   A flow diagram showing the sequence of operations which were used with the MCMC search and upper limit estimation codes.
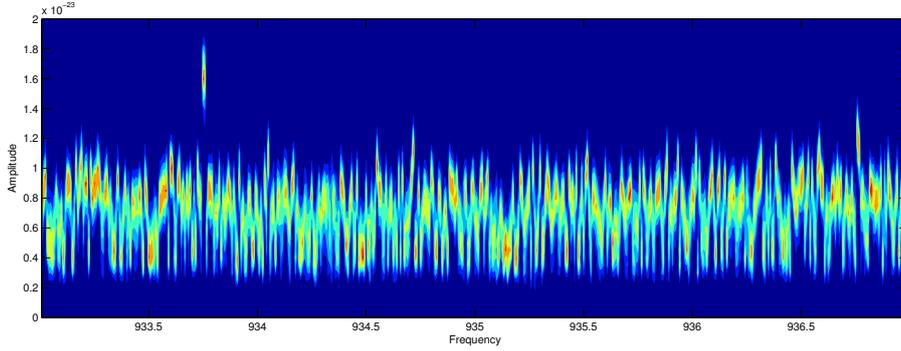
Figure 4.10: The posterior PDFs for each search window marginalised onto the h_0 axis showing an consistent level of noise across the band at around 7×10^{-24}, except at the elevated point shown in figure 4.11.

4.3.3 Results from Search for a Pulsar in SN1987A Data

Searches

Each of the 479 individual frequency bands were searched in parallel using the Caltech computing cluster, with each Markov chain started at a point in parameter space randomly selected from the prior range. None of these searches returned a chain which passed the convergence criteria for the automatic detection of a signal. From the results observed in 3.1.1 this would lead us to expect a limit that the amplitude parameter would settle on, dependent on the noise level. Plotted in figure 4.10 are the marginal distributions for amplitude in each of the 479 search bands, which reveals that there is a consistent level of noise at which the chains arrive, due to the inevitable periodicities which are found when searching over a range of frequencies.

There is an interesting anomaly in bands 90 and 91 which are centred at $f = 933.75\,\mathrm{Hz}$ and $f = 933.75833\,\mathrm{Hz}$, where the marginal amplitude distributions are

elevated clear of the noise floor, as in figure 4.11. If we examine the posterior PDF for the bands we see that the chains cannot be deemed to have converged by the criteria expressed above, but there is a definite sinusoidal component in the noise at this frequency.

Taking a closer look at the data from these frequency bands reveals the cause of the problem. In figure 4.12 is shown the amplitude spectral density for the longest section of uninterrupted data from bands 90 and 91 (490 minutes), with the frequency on the x axis showing the overlapping region between the two. It can easily be seen that there is an excitation of the noise at 933.754 Hz with a maximum amplitude of around 1.6×10^{-23}. This line in the spectrum is clearly the cause of the unusually high marginal amplitude distribution seen in these bands. This means that the probability distribution for these two bands both contain a maximum which is further from zero than in the surrounding frequency bands, and the algorithm has located this maximum and attempted to fit a pulsar signal to it. The MCMC algorithm even produces a reasonably accurate estimation of the amplitude of the line by strongly favouring the model with $\cos \iota \approx 0$, thereby placing all the power in the $+$ polarisation. The cause of the line itself remains unknown.

The outstanding feature in the spectrum of estimates having been dismissed, the search has failed to discover a gravitational wave signal in the frequency region of interest. In accordance with the pipeline, attention was turned to setting an upper limit on the possible amplitude of gravitational radiation that may have been overlooked by the search.
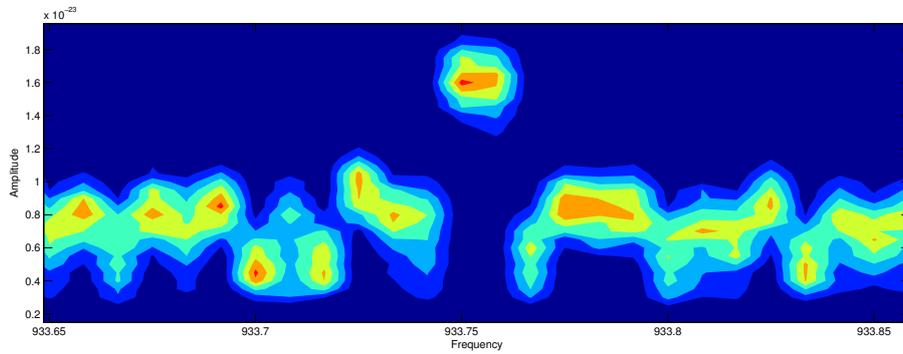
Figure 4.11: Enlarged view of the elevated estimation in figure 4.10 which, although not converged on a signal by the search criteria has an amplitude estimate of around 1.6×10^{-23}. The cause of this elevation is revealed by examining the spectrum of the datasets which produced these chains, shown in figure 4.12.
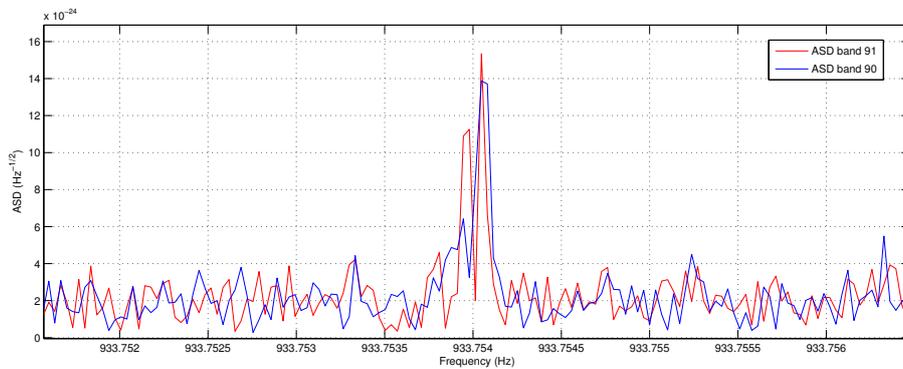
Figure 4.12: The amplitude spectral densities from bands 90 and 91, both showing the anomalous line which lies in the overlapping region.

4.3.4 Upper Limits on Radiation Amplitude in a 4 Hz Band

The Monte Carlo stage used to extract an upper limit from the data and algorithm is here very similar to that used in section 3.1.3 on the artificially generated white noise. There are minor differences which should be mentioned, however, in that the grid used to evaluate the probability distribution was rectangular, extending over 11 regularly spaced points in the range $\cos \iota \in [0, 1]$, and 17 points in the h_0 direction ranging from 1×10^{-23} to 9×10^{-23} so as to fully encompass the range of values in the $(h_0, \cos \iota)$ plane over which the detection probability varies. At each of these points, $M = 36$ injections were made into data taken from the central band at 935 Hz, with noise as estimated from the 4 Hz band.

At each point a frequentist probability of overlooking the signal is evaluated as $P(\text{non-detection}|h_0, \cos \iota) = N_{\text{non-detection}}/M$, defining a 2 dimensional probability function shown in figure 4.13.

This matrix is marginalised onto the h_0 axis, where the upper limit is shown where the distribution falls to the threshold value 0.05 with

$$P(\text{detection}|h_{95\%}) = 1 - P(\text{non-detection}|h_{95\%}) = 0.95,$$

giving an upper limit of $h_{95\%} = 7.3 \times 10^{-23}$, when using linear interpolation between the points of the grid.

This is comparable with the limit reported by the TAMA group in [46] of 5×10^{-23}, which was achieved with an exhaustive matched filtering technique, but with a smaller parameter range of 0.1Hz in δf and 1×10^{-10} in $\dot{\delta f}$. With the
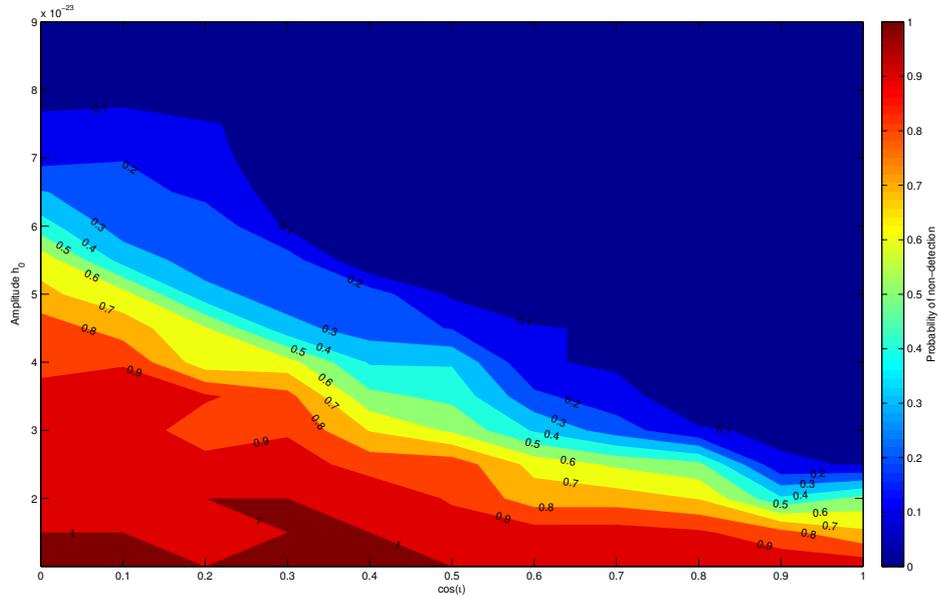
Figure 4.13: The probability distribution shown on the $(h_0, \cos\iota)$ plane for Monte Carlo injections into H1 S3 data at 935 Hz.
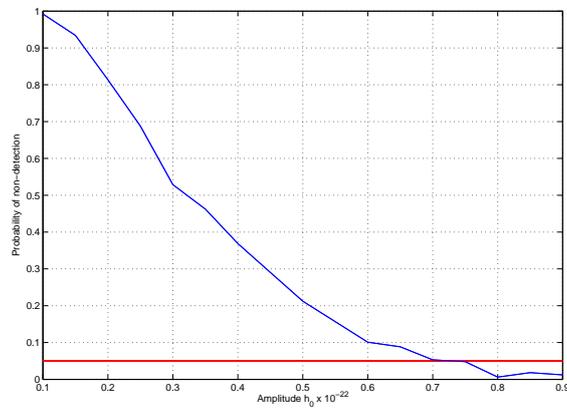
Figure 4.14: The marginalisation of figure 4.13 onto the h_0 axis, showing the decreasing probability of overlooking a signal as the amplitude increases. The 95% upper limit is marked where the distribution in blue has value 0.05 shown as the red horizontal line, at $h_{95\%} = 7.3 \times 10^{-23}$.

more sensitive LIGO data collected since the S3 run it is likely this limit would be exceeded by an MCMC search which could be performed in future using the pipeline I have described.

4.4 Comments

4.4.1 Computational Cost

One of the major justifications for the MCMC technique is the advantages it offers in terms of computational efficiency over the other search methods. To quantify this I will attempt to estimate the cost of running the SN1987A search using the F-statistic method and compare it with the recorded figure for the MCMC search.

A direct comparison is difficult to make as the MCMC operates with a continuous parameter space whereas a search such as the $\mathcal{F}$-statistic [20] uses a discrete grid of points at which $2\mathcal{F}$ values which quantify significance are computed. In practice one would never use an $\mathcal{F}$-statistic search that has less than one template per Fourier bin of the data, but this provides a sensitivity greater than that from the probabilistic MCMC algorithm, which is quantified with the constant $k = 494$ in sensitivity equation $h_{95\%} = k\sqrt{S_h/T_{\text{obs}}}$ from the SN1987A upper limit.

For this $\mathcal{F}$-statistic search the number of templates is calculated assuming the use of a rectangular grid over the frequency and spindown parameters and a minimal overlap of one template per frequency bin in the f direction. The number of templates in the $\dot{f}$ direction can be estimated with the derivation of the phase mismatch as a function of template separation in the $\dot{f}$ direction. From this one finds the result that in order to have a template every π radians (an absolute

minimum number) with the search information described above of $2.82 \times 10^{-14} \Delta \dot{f}$. These values multiplied as in a rectangular grid gives $N_T = 1.7026 \times 10^{11}$ templates that must be searched.

When executed on a 2 GHz cluster node, which is a typical specification for clusters available at present, it takes approximately 20 milliseconds to compute one value of $2\mathscr{F}$ for a dataset of 69 days using 1800 second long SFTs, according to [47]. This would give a total search time of 944 000 CPU-hours on a 2 GHz machine.

To perform the equivalent search with the MCMC took an average of 12 hours per chain on 1.8GHz compute nodes of the Caltech cluster. Over 479 chains this sums to give 5748 CPU-hours for the search stage, and an additional 80,784 CPU-hours to calculate the upper limit from Monte Carlo injections. If we make the approximation that to scale the $\mathcal{F}$-statistic to the same CPU speed we multiply the time taken by (2GHz/1.8GHz), then we arrive at a figure of 1 050 000 CPU hours to perform the search on an equivalent computing system.

From this we can see that the MCMC does indeed offer considerable benefits over the $\mathcal{F}$-statistic search in terms of reducing computation time by a factor of 200 purely in the search regime in this particular example. When it comes to setting upper limits, there will be a similar speedup when comparing equal numbers of injections in each search type, although in the $\mathcal{F}$-statistic search h_0 and $\cos \iota$ do not appear to be parameters that are varied in setting upper limits, so a like-for-like comparison is not drawn.

It must however be clear that the sensitivity of the MCMC search is not as great as that of the $\mathcal{F}$-statistic, for the reason already discussed that in a particular search band, the MCMC chain will home in on the frequency component that has

maximum amplitude and neglect those with lower noise levels. The $\mathcal{F}$-statistic search on the other hand examines each frequency bin exhaustively within the band, effectively sampling more templates that lie at lower noise levels and allowing a distribution to be built up from the entire space of $\delta f, \dot{\delta f}$ parameters. It remains true however that when searching over a large range of frequencies the sensitivity of the $\mathcal{F}$-statistic is reduced for the same basic reasons, and the study of this effect in [48] suggests that for a dataset with 6×10^6 seconds of data the value of k will increase to between 30 and 40, degrading the sensitivity of the search accordingly.

This level is still more sensitive than the MCMC search by an order of magnitude, however. This suggests that a Markov chain Monte Carlo algorithm search is less suited than the traditional $\mathcal{F}$-statistic search for the current generation of detectors, where setting upper limits is the typical application. However it may find an ideal application as the sensitivity of detectors brings pulsar signals into the SNR range that would make them detectable by the MCMC search. In such a case the algorithm would provide a fast way to quickly scan for and perform parameter estimation on loud signals, without the need for massively distributed computing platforms such as the Einstein@home search effort which distributes $\mathcal{F}$-statistic searches to thousands of desktop computers whose CPU cycles are donated by volunteers worldwide. It must still be said, however, that an $\mathcal{F}$-statistic search would still be better suited to the detection of weak signals.

4.4.2   Conclusions Drawn from This Work

Based on the results described in this chapter, I conclude that the use of this MCMC technique is not the optimal means of searching for very weak amplitude

continuous signals, where the noise dominates the data. The increased efficiency in computational cost by a factor of approximately 200 has to be weighed against the reduced sensitivity of the search. In the current regime, where an initial detection is the goal of most search programmes, sensitivity is the key quality which distinguishes the performance of an algorithm and in these terms this MCMC implementation does not compare well.

The reason for this behaviour at low SNR is, as stated above, that the efficiency gains of MCMC depends on the structure of the posterior in parameter space. In situations where there is a gradient of probability that the MCMC can follow, it does not need to explore the entire space to find the maximum. This is the case when the signal is strong, and triggers raised probability in nearby frequency bins, giving the algorithm a good chance of finding the signal. However, when the signal is weak, the parameter space is dominated by noise, and especially in the frequency - spindown plane there is no information about the maximum for the algorithm to follow. In this case, there is no benefit to using a probabilistic search, as an exhaustive search will find the signal, and will do so in less time, as MCMC will visit the same state more than once whereas a grid-based search simply proceeds through the space without revisiting any one point.

One may speculate, however, that when future generations of detectors come online with vastly improved sensitivity, searches which require the exploration of a very large parameter space will be more amenable to techniques such as the one I have developed. In such a situation it would be far more efficient to perform a directed MCMC search which can return a full Bayesian estimation of the parameters of the source without having to explore the parameter space exhaustively. Indeed, in any situation where the signal to noise ratio makes the application of

an MCMC or other probabilistic algorithm a possibility, the increased efficiency of such techniques should ensure that they are favoured over exhaustive methods. This is already apparent in certain areas of cosmology where MCMC has found many applications, and in other statistical sciences such as biology it has long been in use.

4.4.3 Extensions of the Algorithm and Future Applicability

At the time of writing, Markov Chain Monte Carlo methods are increasingly being considered as the core of gravitational wave searches in both ground- and space-based detectors.

With the success of the MCMC algorithm in decreasing processing time for the six parameters search, an extension to cover 8 parameters with the addition of position of the source on the sky parametrised by ecliptic latitude and longitude has been developed by Richard Umstätter in collaboration with Nelson Christensen. This increases further the speed benefit given by MCMC as these extra search parameters impose a greater penalty on an exhaustive search than on an MCMC search.

Work is also underway by Chris Messenger to produce a frequency-domain search for pulsars in binary systems based on MCMC which utilises the efficiency of the Fast Fourier Transform in combination with Markov chain Monte Carlo techniques to explore a larger parameter space where the orbital parameters of the binary system are also estimated by the routine.

Alternatively, the MCMC method could be used as part of a pipeline, to per-

form parameter estimation on candidate signals that are identified by another search. In this usage the MCMC would provide a tool for extracting astrophysical information, with the actual search being performed by an exhaustive method.

It is likely that this technique will continue to be honed by the data analysis community for use in gravitational wave analyses of many types, and the use of a Bayesian framework allows a natural extension to multi-detector network analyses as collaborations are built combining data from the LIGO, GEO and VIRGO projects.

I will now turn to another application of MCMC in the proposed space-based detector LISA, where continuous waves from binary systems are the sources as opposed to those of a rotating neutron star.

Chapter 5

An MCMC Approach to Characterising Galactic Binary Systems in LISA

In this chapter I will move on to discussing work on data analysis for the planned Laser Interferometer Space Antenna (LISA), in which the techniques of MCMC have again been applied to a data analysis problem of a rather different type. I will first discuss the nature of the problem, then outline the approach to solving the problem by gradually increasing the complexity of the signals under consideration in sections 5.3 and 5.4.

5.1 The LISA Binary Source Confusion Problem

LISA is a mission which is under joint development by the European Space Agency and NASA which aims to complement the already existing ground-based detectors

by launching a triplet of spacecraft which will act as a 3-sided laser interferometer, with an arm length of approximately 5 million kilometres. The advantages of placing a gravitational wave antenna in space are twofold. Firstly, as they are in free-fall, the end test masses of the interferometer will be completely isolated from all seismic noise, and being in space it is far removed from the sources of gravity gradient noise which prevent terrestrial detectors from operating at frequencies below around 10 Hz.

Secondly, the enormous arm length of the space based interferometer means that the passing gravitational waves will produce a proportionally greater strain on the proper distance between the test masses, as in equation 1.7, and cause a larger phase shift in the laser. This means that to reach an equivalent strain sensitivity as a ground-based detector requires less precise measurement of the phase of the incoming light at the vertices of the triangle. However, the technology required for the construction of LISA is itself more challenging.

With an arm-length of 16.6782 light-seconds, approximately 5 Gm, LISA will have a sensitivity to gravitational waves with wavelengths considerably longer than those available to LIGO. Specifically, it is designed to have optimal sensitivity in the frequency range 0.01 mHz to 100 mHz [49], which will allow the observation of a great many astronomical sources in a completely different part of the gravitational spectrum, as compared to ground-based detectors. These sources are mentioned in section 1.3, and in particular LISA is expected to observe gravitational waves emitted by the merging of supermassive black holes at cosmological distances, the last orbits of compact objects around massive black holes and a possible gravitational wave stochastic background.

In this frequency band there are also expected to be sources within the Milky

Way, in the form of binary star systems whose quadrupole motion will emit gravitational waves at twice their orbital frequencies, as in section 1.2.3. The galactic population of such systems, which are compact binaries containing white dwarf stars or neutron stars is estimated to be up to 100 000 in the 1-5 mHz band [13]. The continuous radiation from these sources will swamp the detector to a level above the instrumental noise curve. Those binaries with an emission frequency less than 1 mHz will be unresolvable, forming a background noise of gravitational waves; above this frequency there will be a great many sources which overlap in frequency, making it a difficult problem of data analysis to separate them. Identifying and characterising these binary systems is the LISA binary source confusion problem, and here I shall describe the approach to tackling this problem that has been developed based again on a Markov chain Monte Carlo approach.

5.2 Occam's Razor

An important aspect of this problem that should be considered is the difficulty in deciding on a suitable number of sources with which to fit the data. It is plain to see that the fit to the data may be made arbitrarily good with an unlimited number of sources being posited, and in fact in the case of sinusoidal signals, the result is simply the Fourier series representation of the noisy data. Since the number of sources in our problem is unknown, we must have a way of knowing when to stop adding sources and overfitting the data. This is a classic example of the problem to which one would apply the metaphorical Occam's Razor, which dictates that the simplest explanation which adequately describes the data is the most preferable, or in Einstein's words, "Everything should be made as simple as possible, but not

simpler."

To be completely accurate it is not, in fact, the complexity of the model per se that should be discriminated against by the principle, but its predictive power, where a more complex hypothesis generally has a broader range of measurements which it can accommodate, whereas a simple hypothesis "risks" more by allowing for a smaller range of possible outcomes. If the result of the experiment is indeed allowed by the simple hypothesis we would instinctively favour it over the less predictive one.

One very useful and appealing feature of Bayesian inference is that it automatically contains this intuitive principle within its formulation in a quantitative way; and can in fact set relative probabilities for competing hypotheses of differing complexity or predictive power. I will illustrate this with a very simple example.

Consider two competing hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$ which are to be evaluated in the light of some observational or experimental data $\{d\}$. The probability of $\mathcal{H}_1$ is $P(\mathcal{H}_1|\{d\}, I) = P(\mathcal{H}_1|I)P(\{d\}|\mathcal{H}_1, I)/P(\{d\}|I)$ from Bayes' Theorem, with a similar expression for $\mathcal{H}_2$. The *odds ratio* $\mathcal{O}_{12}$ of these two hypotheses, which is a factor indicating the probability of one in relation to the other, is given by dividing the individual model probabilities, where the term $P(\{d\}|I)$ cancels.

$$\mathcal{O}_{12} = \frac{P(\mathcal{H}_1|\{d\}, I)}{P(\mathcal{H}_2|\{d\}, I)} = \frac{P(\mathcal{H}_1|I)}{P(\mathcal{H}_2|I)} \frac{P(\{d\}|\mathcal{H}_1, I)}{P(\{d\}|\mathcal{H}_2, I)}. \tag{5.1}$$

We will assume that we have no prior information to favour one model over the other, therefore $P(\mathcal{H}_1|I)/P(\mathcal{H}_2|I) = 1$, and we are only concerned with the ratio of likelihoods.

For the sake of argument, let the two hypotheses be dependent on a single

parameter x, where the difference between the two is the range that x is allowed to take: $x_1 \in (0,1)$, $x_2 \in (0,2)$. That is to say that $\mathcal{H}_1$ has greater predictive power than $\mathcal{H}_2$, or that it is more precise in its predictions. Applying the product and sum rules of probabilities we have,

$$P(\{d\}|\mathcal{H}_1|I) = \int_0^1 p(\{d\}|x_1, \mathcal{H}_1, I)p(x_1|\mathcal{H}_1, I)dx_1, \qquad (5.2)$$

and

$$P(\{d\}|\mathcal{H}_2|I) = \int_0^2 p(\{d\}|x_2, \mathcal{H}_2, I)p(x_2|\mathcal{H}_2, I)dx_2. \qquad (5.3)$$

We shall assume the case of no noise and a single, exact observation d so that $p(d|x_1, \mathcal{H}_1, I) = p(d|x_2, \mathcal{H}_2, I) = \delta(d - x_{\{1,2\}})$ (where $\delta()$ is the Dirac delta function), lending equal support to either hypothesis. The priors are assumed to be flat, and must be normalised such that $p(x_1|\mathcal{H}_1, I) = 1$ and $p(x_2|\mathcal{H}_2, I) = \frac{1}{2}$. We are now ready to write down the expansion of the odds ratio when an observation is made,

$$\mathcal{O}_{12} = \frac{1 \times \int_0^1 \delta(d - x_1)dx_1}{0.5 \times \int_0^2 \delta(d - x_2)dx_2} = \begin{cases} 2 & 0 \le d \le 1 \\ 0 & 1 < d \le 2 \end{cases} \qquad (5.4)$$

This shows that as one would expect, the more predictive model is favoured if an observation that falls into its prior range ($0 \le d \le 1$), but if the observation contradicts the prediction of the hypothesis, then its probability vanishes as the data rules it out. Although this is a simple example, it is illustrative of the broader principle that is automatically applied when using Bayesian inference in this way. We can see that the reason that simpler hypotheses are favoured while the data does not rule them out, is that they spread their prior probability density

over a smaller range. A slightly more developed example where hypotheses with different numbers of parameters are compared is given in [2]. By following this approach to the problem we automatically incorporate selection of the simplest model explaining the data.

5.3   Detecting multiple sinusoids in noise

5.3.1   Description of model

The strategy that was adopted was to approach the problem using a toy model, and add complexity to work towards simulating the full problem of the realistic analysis. The initial iteration of the analysis was a MCMC code which was developed by Richard Umstätter to estimate the number of signals in noisy data and their parameters, where the individual signal waveform was a simple sinusoid with three parameters, two amplitudes A, B and a frequency $\omega = 2\pi f$,

$$s(t_j; \mathbf{a}) = A\cos(\omega t_j) + B\sin(\omega t_j), \tag{5.5}$$

which can also be expressed as $s(t_j; \mathbf{a}) = h\cos(\omega t_j - \phi_0)$ where $h = \sqrt{(A^2 + B^2)}$ and $\phi_0 = \tan^{-1}\frac{B}{A}$. The combined signal from m sources is simply given as the summation of these,

$$y(t_j; \mathbf{a}_m) = \sum_{i=1}^{m} A_i\cos(\omega_i t_j) + B_i\sin(\omega_i t_j). \tag{5.6}$$

For a model $\mathcal{M}_m$ which contains m signals, we therefore have a $(3m + 1)$-

dimensional parameter vector

$$\mathbf{a}_m = [A_1^{(m)}, B_1^{(m)}, f_1^{(m)}, A_2^{(m)}, B_2^{(m)}, f_2^{(m)}, \dots, A_m^{(m)}, B_m^{(m)}, f_m^{(m)}, \sigma_m^2] \tag{5.7}$$

where the noise level σ^2 is included as a parameter and is assumed to be stationary throughout the dataset, with likelihood function

$$p(\{d\}|\mathbf{a}_m, I) \propto \frac{1}{\sigma_m^N} \exp\left[-\frac{1}{2\sigma_m^2} \sum_{j=1}^N (d_j - y(t_j; \mathbf{a}_m))^2 \right] \tag{5.8}$$

An alternative way of thinking about the problem of selecting between the models would be consider m itself a parameter which will be estimated by the code along with the individual parameters of each gravitational wave.

For simplicity, flat priors are used for the noise level σ_m^2 and each amplitude parameter A and B, although it should be noted this produces a different distribution from a flat prior in polar co-ordinates h and ϕ_0, the posterior distribution is not significantly affected by this choice when the data contains detectable signals.

The parameter space is then explored by a MCMC sampler, as described in [2] and [50] by Umstätter *et al*, which implements the Reversible Jump MCMC algorithm first described in [1] by Peter Green to make transdimensional jumps. I shall here briefly cover the means by which these jumps are made, a more detailed description is given in the references. The implementation also uses a delayed rejection stage similar to that used in the pulsar code, and which has already been introduced in 2.5.2. This algorithm, originally implemented by Umstätter *et al* was the basis for my work on approaching the full LISA problem which begins with section 5.4.

5.3.2 Transdimensional Jumps

If at iteration n we hold k signals in the model, parametrised by $\mathbf{a}_k$, and we want to move into a state that has either $k+1$ or $k-1$ signals. In order to preserve the principle of detailed balance, the transdimensional move must match the number of dimensions between models. This is achieved by the random generation of a parameter triple $\mathbf{r}$ from a proposal distribution $q(\mathbf{r})$, and the probabilities are equal for moving up and down between dimensionalities, i.e. $p_{k \mapsto k'} = p_{k' \mapsto k}$.

Four types of move are available for proposal within the algorithm to move from a model of one dimensionality to another of higher or lower dimensionality (except when the model is in the state $m = 0$ from where the dimensionality can only increase. These are proposed at random with probability $P = 0.3$ at any iteration, and if this condition is met they replace a intra-dimensional move for that iteration.

Birth and death

The birth move extends the dimensionality of the model by simply proposing the addition of a new signal with parameters $\mathbf{a}'_{(i)}$ into the model with parameters chosen at random, independent of the existing signals in the model $\mathcal{M}_k$. The prior probability of the k-th model is denoted $p(k)$, but these prior probabilities are assumed to be equal in this implementation. The acceptance probability given is by [2] as

$$\alpha_{k \mapsto k+1}(\mathbf{a}'_{k+1}|\mathbf{a}_k) = \min \left\{ 1, \frac{p(k+1)p(\mathbf{a}_{(i)})p(\mathbf{d}|\mathbf{a}'_{k+1}, k+1)}{p(k)p(\mathbf{d}|\mathbf{a}_k, k)q(\mathbf{r})} \right\}, \qquad (5.9)$$

In the proposal distribution $q(\mathbf{r})$, the frequency of the new signal is chosen by randomly generating a frequency between 0 and f_{Nyquist} from a uniform distribution, and comparing the power of this new signal as calculated from the periodogram

$$C(f) = \frac{1}{N}\left[\sum_{i=1}^{N} d_i^2 \cos^2(2\pi f t_i) + \sum_{i=1}^{N} d_i^2 \sin^2(2\pi f t_i)\right],\qquad(5.10)$$

with a random number generated uniformly between 0 and the maximum power in any frequency bin which is precomputed using a fast Fourier transform during an initial pass over the data. Amplitude parameters are chosen from a normal distribution $N(0, \sigma)$ with mean zero and σ given by the mean amplitude of the signals already present in the model.

The inverse death transformation chooses a signal at random from those already present and eliminates it from the model, reducing the parameter space by 3 dimensions. The acceptance probability is

$$\alpha_{k\mapsto k-1}(\mathbf{a}'_{k-1}|\mathbf{a}_k) = \min\left\{1, \frac{p(k)p(\mathbf{d}|\mathbf{a}_k, k)q(\mathbf{a}'_{(i)})}{p(k-1)p(\mathbf{a}'_{(i)})p(\mathbf{d}|\mathbf{a}'_{k-1}, k-1)}\right\}.\qquad(5.11)$$

Both these transitions have a Jacobian determinant $|J_{k\mapsto k'}| = 1$.

Split and merge

The split transition generates an increase in the number of signals by picking an existing signal and splitting it into two resultant signals, each perturbed by a small random vector $\mathbf{r}$ which is drawn from a 3 dimensional Gaussian distribution with mean zero. The amplitudes are halved during the transition so the combined

amplitude is conserved. The transition of signals is given by the equation,

$$
t_{k \mapsto k+1}(\mathbf{a}_{(i)}, \mathbf{r}) =
\begin{pmatrix}
\frac{1}{2} A_i^{(k)} + r_A \\
\frac{1}{2} B_i^{(k)} + r_B \\
f_i^{(k)} + r_f \\
\frac{1}{2} A_i^{(k)} - r_A \\
\frac{1}{2} B_i^{(k)} - r_B \\
f_i^{(k)} - r_f
\end{pmatrix}
=
\begin{pmatrix}
A_{i_1}^{(k+1)} \\
B_{i_1}^{(k+1)} \\
f_{i_1}^{(k+1)} \\
A_{i_2}^{(k+1)} \\
B_{i_2}^{(k+1)} \\
f_{i_2}^{(k+1)}
\end{pmatrix}.
\tag{5.12}
$$

The acceptance ratio of this transition is

$$
\alpha_{k \mapsto k+1}(\mathbf{a}'_{k+1} | \mathbf{a}_k) = \min \left\{ 1, \frac{p(k=1, \mathbf{a}'_{(i_1)}, \mathbf{a}'_{(i_2)}) p(\mathbf{d} | \mathbf{a}'_{(i_1)}, \mathbf{a}'_{(i_2)}, k+1)}{p(k, \mathbf{a}_{(i)}) p(\mathbf{d} | \mathbf{a}_{(i)}, k) q(\mathbf{r})} \right\} |J_{k \mapsto k+1}|,
\tag{5.13}
$$

with Jacobian $|J_{k \mapsto k+1}| = 2$.

The inverse merge process picks two sinusoids at random and proposes that they be combined with the transformation

$$
t_{k \mapsto k-1}(\mathbf{a}'_{(i_1)}, \mathbf{a}'_{(i_2)}) =
\begin{pmatrix}
A_{i_1}^{(k)} + A_{i_2}^{(k)} \\
B_{i_1}^{(k)} + B_{i_2}^{(k)} \\
\frac{1}{2}[f_{i_1}^{(k)} + f_{i_2}^{(k)}] \\
\frac{1}{2}[A_{i_1}^{(k)} - A_{i_2}^{(k)}] \\
\frac{1}{2}[B_{i_1}^{(k)} - B_{i_2}^{(k)}] \\
\frac{1}{2}[f_{i_1}^{(k)} - f_{i_2}^{(k)}]
\end{pmatrix}
=
\begin{pmatrix}
A_i^{(k-1)} \\
B_i^{(k-1)} \\
f_i^{(k-1)} \\
r_A^{(k)} \\
r_B^{(k)} \\
r_f^{(k)}
\end{pmatrix},
\tag{5.14}
$$

where $\mathbf{r} = (r_A, r_B, r_f)$ is a vector of the half-distances between the merged parameters, which is then used to calculate the proposal probability $q(\mathbf{r})$, which is used

in the acceptance ratio:

$$\alpha_{k \mapsto k-1}(\mathbf{a}'_{k-1}|\mathbf{a}'_k) = \min\left\{1, \frac{p(k-1, \mathbf{a}'_{(i_1)}, \mathbf{a}'_{(1_2)})p(\mathbf{d}|k-1, \mathbf{a}'_{(i_1)}, \mathbf{a}'_{(1_2)})}{p(k, \mathbf{a}_{(i)})p(\mathbf{d}|k, \mathbf{a}_{(i)})q(\mathbf{r})}\right\}|J_{k \mapsto k-1}|. \tag{5.15}$$

The Jacobian here is the inverse of that of the split, therefore $|J_{k \mapsto k-1}| = 1/2$.

The proposal distribution $q(\mathbf{r})$ is a multivariate normal distribution with mean zero and diagonal covariance matrix $\text{diag}(\sigma_A, \sigma_B, \sigma_f)$, where $\sigma_A = \sigma_B = \sigma_m$, the current estimated noise level in the model. σ_f is calculated based on the work in [39], where the width of the frequency posterior mode is a function of the SNR and the number of data points N and is given by

$$\sigma_f = \left(2\pi\sqrt{\frac{(A^2 + B^2)}{\sigma_m^2}}\right)^{-1}\sqrt{\frac{48}{N^3}} \tag{5.16}$$

This proposal distribution for frequency is also used when proposing jumps in that parameter without changing the dimensionality of the model.

5.3.3 Results

In tests of the code, 100 signals with random amplitudes A and B in the range [-1,1] were injected into a dataset of length 1000 datapoints consisting of white Gaussian noise of variance $\sigma = 1$. The results of this simulation are described in [2] in great detail. I will simply report here that the code produced the greatest probability for the presence of 95 signals and a noise level of $\sigma = 1.025$, the sinusoids which were not recovered contributing their power to the noise estimate instead. This reflects the desired behaviour for the algorithm in approaching the toy problem, as in the real LISA data analysis there will be some binaries that are not resolvable in

the data, and in the absence of the ability to recreate their signal, it must simply be allocated as noise.

5.4 Extending the Algorithm With Source Location Parameters

The work from this section onward was based on the algorithm developed by Umstätter *et al*, but was undertaken by myself alone. I will now describe the modifications I have added to the above model in order to allow sky position to be estimated along with frequency amplitude and phase. This represents a small step toward the full LISA problem: in reality LISA will have multiple data streams based on time-delay interferometry, which have complicated response function themselves, and are further complicated by amplitude and frequency modulation as LISA moves in its orbit [51],[52]. In the model adopted here the signal is modulated in frequency by the orbit of LISA, but antenna itself is assumed to be isotropic and report only one channel of the signal, so polarisation is not considered.

The new form of the signal is now a function of 5 parameters, with the addition of λ the ecliptic longitude and β the ecliptic latitude. The prior on the longitude parameter is flat and constant $p(\lambda, I) = \frac{1}{2\pi}$ and its range $\lambda \in [0, 2\pi)$. The latitude β takes on a non-uniform prior given by $p(\beta, I) = \frac{1}{2}\cos\beta$ in the range $\beta \in [-\frac{\pi}{2}, \frac{\pi}{2})$, as the two co-ordinates together span the surface of the celestial sphere.

The phase of the sinusoidal signal as it arrives at the detector is then modulated

in the following way, as LISA moves toward and away from the source:

$$s(t_j; \mathbf{a}) = A \cos\left(\omega\left[t_j - \Phi_D(t_j; \mathbf{a})\right] - \phi_0\right) \tag{5.17}$$

where $\Phi_D(t_j; \mathbf{a})$ is the phase modulation given by

$$\Phi_D(t_j; \mathbf{a}) = \frac{\mathbf{n}.\mathbf{r}}{c} = \frac{R}{c}\cos\beta\cos(\lambda - \Omega t_j). \tag{5.18}$$

With $\mathbf{n}$ being a normal vector directed from the SSB to the position of the source on the sky, and $\mathbf{r}$ the vector from the SSB to the position of LISA's centre. Here the radius of LISA's orbit in light seconds is $R = 499.004\,\text{l.s.}$ and $\Omega = 2\pi \times (1\text{yr})^{-1}$ is the angular frequency of the orbit, which is assumed here to be circular and have a radius of 1 astronomical unit.

5.4.1 Implementation

With the addition of the two new parameters, suitable proposal distributions had to be included with them, and the transdimensional steps updated also. In particular, the merge routine had to be altered such that the new signal which is created by the fusion of the two chosen ones had a position on the sky which lay halfway between the originating signals. Since there are two halfway points on the sphere, the one in the middle of the shorter arc segment is chosen. The procedure for calculating this point finds the vector average $\frac{1}{2}(\mathbf{n}_1 + \mathbf{n}_2)$ of the two normal vectors in three dimensions $\mathbf{n}_i = (\cos\beta_i\cos\lambda_i, \sin\lambda_i\cos\beta_i, \sin\beta_i)$, $(i = \{1, 2\})$, and extending the resulting vector to find its intersection with the unit sphere. The new co-ordinates $\beta_i^{(k-1)}$ and $\lambda_i^{(k-1)}$ are given here explicitly as functions of the

spherical co-ordinates of the originating signals labelled 1 and 2 as,

$$\cos\theta = \cos\beta_1\cos\beta_2\left[\sin\lambda_1\sin\lambda_2 + \cos\lambda_1\cos\lambda_2\right] + \sin\beta_1\sin\beta_2 \quad (5.19)$$

$$\beta_i^{(k-1)} = \sin^{-1}\frac{\sin\beta_1 + \sin\beta_2}{2\cos\left(\frac{\theta}{2}\right)} \quad (5.20)$$

$$\lambda_i^{(k-1)} = \pi + \tan^{-1}\left(\frac{\cos\beta_1\sin\lambda_1 + \cos\beta_2\sin\lambda_2}{\cos\beta_1\cos\lambda_1 + \cos\beta_2\cos\lambda_2}\right). \quad (5.21)$$

The initial implementation used a uniform random distribution over the whole sky, multiplied by the prior on β to make proposals. This was also used when proposing a new signal in the event of a birth move. A bi-variate Gaussian distribution was used in the split and merge moves for sky position to determine the probability of the signals merging or picking the perturbation of the signal positions if they are splitting. At first this was based on a crude formula which approximated the orbit of LISA as an aperture, but this was improved upon with the work described in 5.4.3. For birth moves a new sky location was chosen from the prior distribution also.

5.4.2 Structure of the Posterior Distribution

After trial runs it became clear that the choice of uniform distributions produced extremely low acceptance ratios in the Markov chain stages which involve steps in sky position after the burn-in stage had passed. To investigate the reason for this, the marginal posterior PDF for sky position was examined in detail. For the purposes of analysing the problem, a code was developed which evaluated the two dimensional posterior PDF slice through the parameter space at the injected values of the amplitude and frequency parameters, when a single signal was injected

into noisy data. This revealed a significant amount of structure present in the distribution, as shown in figures 5.1, 5.2.

There is clearly a symmetry between the north and south hemispheres, as a point at latitude $+\beta$ will produce the same Doppler modulation as one at $-\beta$. Therefore the distributions are symmetric upon reflection at the equator. This degeneracy is broken for the real LISA system as the orientation of the array and the amplitude response function cause the signal modulations to differ between north and south in general.

There is also a ring-like structure which occurs in each case, where the maximum probability is found in a mode which shows structure encircling the acceleration vector of the detector. Over the course of a year's observations these patterns are integrated over the entire sky, destructively interfering for all points but the main probability mode around the true location of the source.

This structure was observed to vary with the frequency of the signal, and with the length of observation. In LISA, the resolvability of a signal is known to increase with its frequency. By comparing figures 5.1 and 5.2, where the frequency of the signal is increased from 1 mHz to 6 mHz, this increased resolution is clearly visible. The low resolvability of signals at low frequency is a contributing factor to the confusion arising at these frequencies, as their posterior PDFs will overlap extensively. By contrast, at high frequencies the accuracy possible becomes much improved, and therefore the posterior PDF is much more localised in one position on the sky.

In addition to this being an important feature of LISA's performance, it will also affect the behaviour of the MCMC code when proposing a change in the estimated position on the sky. At low frequencies a proposal distribution with a large
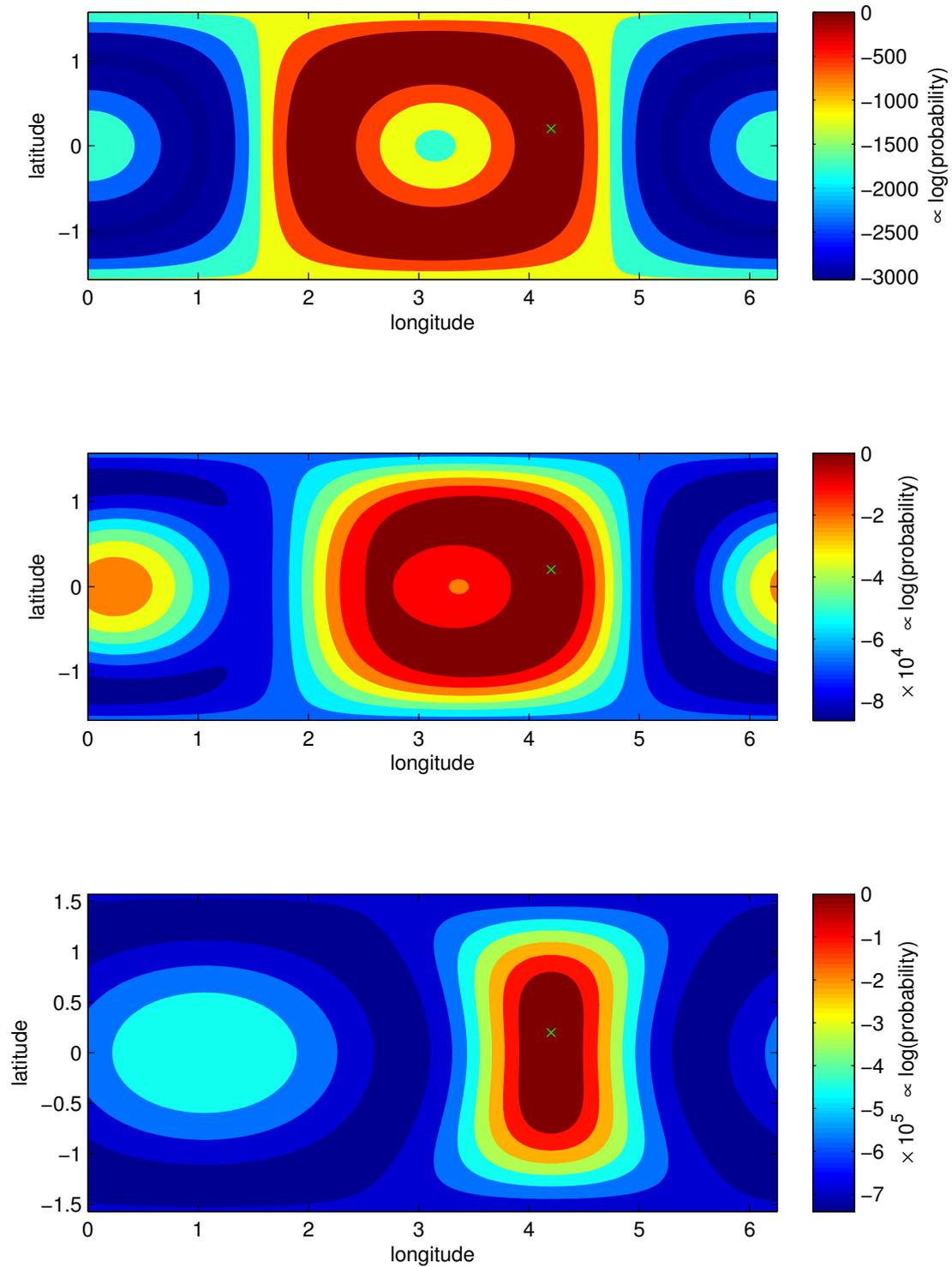
Figure 5.1: The posterior log PDF as calculated as a slice through the sky in ecliptic coordinates, with the other parameters maximised for a signal of frequency $f = 1\,\mathrm{mHz}$ showing the increase in resolution with observation time from 1 day (top), 30 days (middle) and 1 year (bottom).
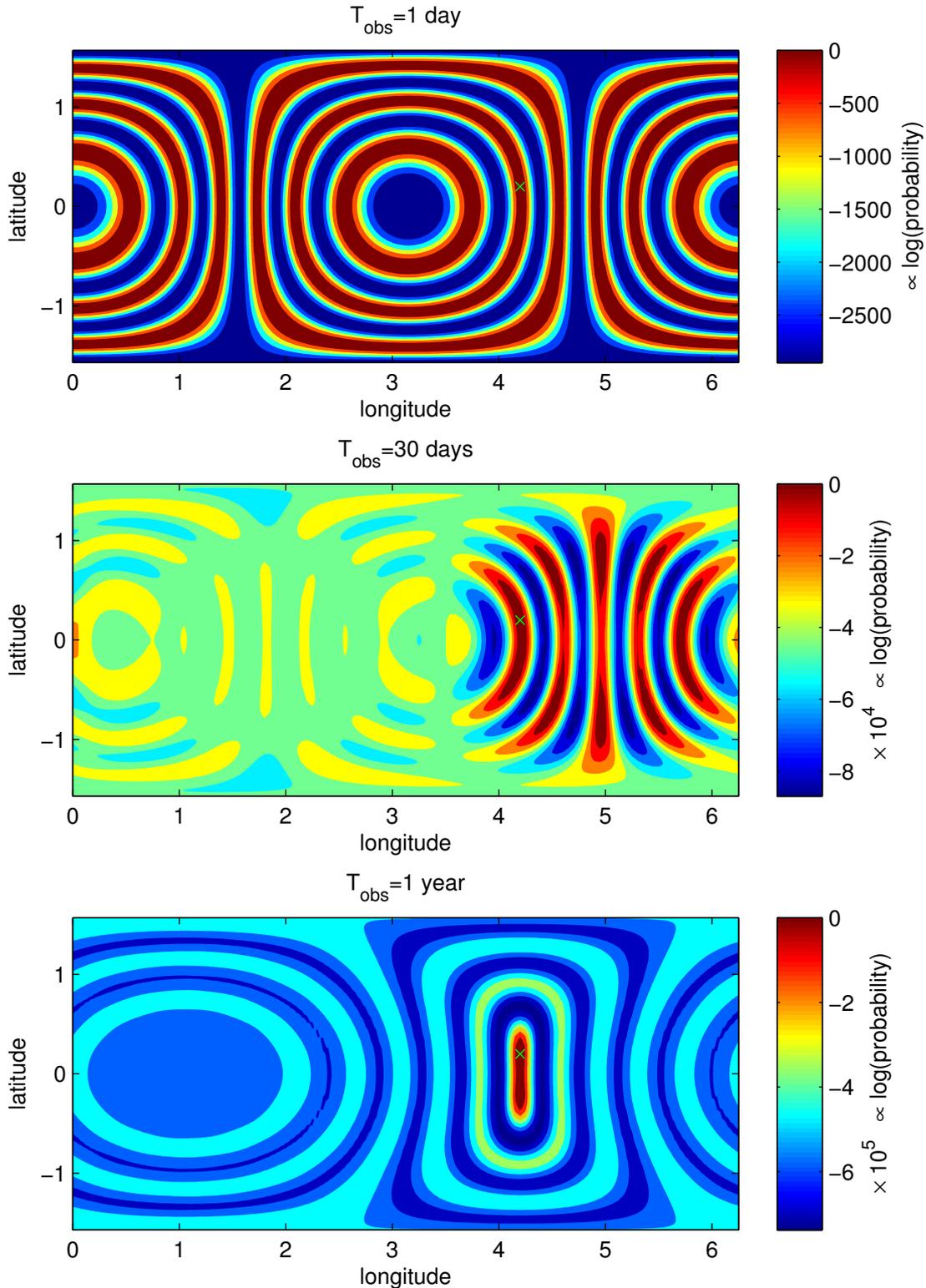
Figure 5.2: As in fig. 5.1, the probability surface for the sky position parameters, but for a signal with frequency $f = 6\,\mathrm{mHz}$. This shows the significantly higher resolvability of a signal at higher frequency, a factor not accounted for in the naïve implementation of the proposal distribution. This plot also reveals the intricate interference pattern centred on LISA's acceleration vector in the observations from 1 day (top diagram) which exhibits $2fR/c$ rings (where R is LISA's orbital radius).

width is appropriate for rapid exploration of the mode, whereas at high frequencies using such a proposal distribution will lead to small acceptance ratios as the majority of proposals will fall well outside the mode in an area of low probability density. This is a factor that should be taken into account in a replacement proposal density to the uniform one described above.

The other factor that is important when considering the proposal distribution is the covariance matrix of the two parameters. The orientation and width of the mode in the sky parameter slice varies not only with frequency of the signal, but also with the position of the source itself, meaning that there are different sensitivities at different positions in the sky. This can be understood intuitively - a source located directly at the ecliptic north pole will exhibit no Doppler shift as the relative motion of LISA along the vector connecting the SSB and source is constantly zero. As the Doppler modulation is a function of $\sin \beta$, its rate of change $\cos \beta$ vanishes at the poles $\beta = \pm \frac{\pi}{2}$, meaning the distinguishability of sources in this region is low, increasing toward the equator. In addition there is a coordinate singularity at the poles, where the longitude is degenerate. This produces a very large width of mode in that direction. More generally the resolvability and orientation of modes varies continuously over the surface of the celestial sphere. Unlike in the previous case of the pulsar search, where a reparametrisation could account for the correlation between parameters, this problem cannot be solved this way; there is an inherent curvature in the parameter space here that cannot be removed with a single transformation that applies at all points.
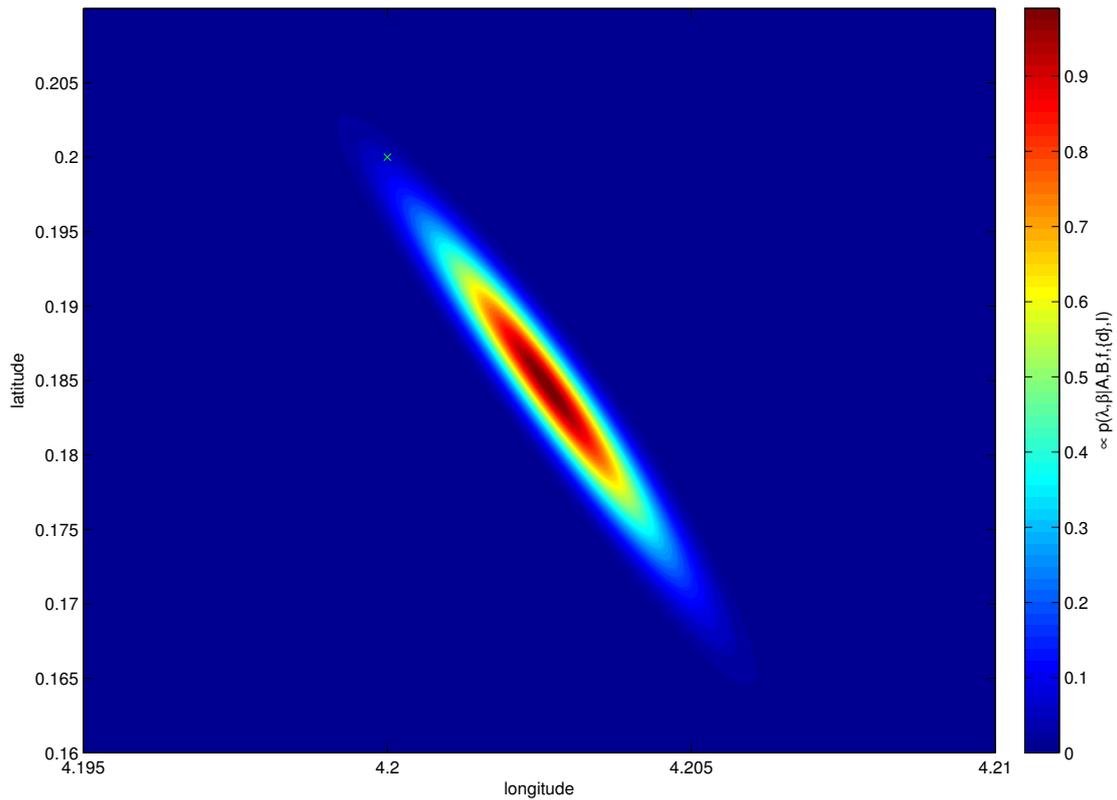
Figure 5.3: A closeup of the posterior probability density function of λ and β with other parameters maximised. It is Gaussian to good approximation and therefore can be described by the covariance matrix approach outlined in 5.4.3. The injected value is marked with a green ×.

5.4.3 A Fisher Matrix Based Approach to Proposal Distributions

Having outlined the problem that faces an MCMC algorithm attempting to make suitable jumps in the posterior parameter space of a single signal, I shall now describe the method which was used to tackle this problem. It will employ an adaptive proposal distribution which is recalculated for each step based on an approximation to the Fisher information matrix which allows a rapid calculation, making the procedure efficient enough to be used at every iteration of the chain.

Instead of a multivariate distribution in 5 dimensions being used at each step, it is simpler from an analytic point of view to divide the steps into three classes and use a different proposal distribution for each. The third type of move is simply added into the algorithm with an equal probability to the existing two types. The three moves then are a change in amplitude A and B; a change in frequency f; or a change in sky position λ and β.

To derive a suitable proposal distribution for the new move type, I have used an approximation based on the Fisher matrix, which will be outlined here. The logarithm of the likelihood function $\log p(\{d\}|\mathbf{a}_i, I) = \mathcal{L}(\mathbf{a})$ can be expressed in the form of a Taylor expansion,

$$\mathcal{L}(\mathbf{a}) = \mathcal{L}(\mathbf{a}_0) + \left.\frac{\partial \mathcal{L}}{\partial \mathbf{a}}\right|_{\mathbf{a}_0} (\mathbf{a} - \mathbf{a}_0) + \frac{1}{2}\left.\frac{\partial^2 \mathcal{L}}{\partial \mathbf{a}^2}\right|_{\mathbf{a}_0} (\mathbf{a} - \mathbf{a}_0)^2 + \dots . \tag{5.22}$$

The step size for a particular mode in the PDF should be on the same order as the width of the mode. Therefore we are interested in knowing the width around the peak of the posterior PDF, at which point the first derivative vanishes $\frac{\partial \mathcal{L}}{\partial \mathbf{a}} = 0$,

and the width of the mode in one dimension is given by the value of the second derivative of the log likelihood.

When considering the two-dimensional case and hold constant A, B and f at their peak values, the relation for λ and β becomes

$$
\begin{aligned}
\mathcal{L}(\lambda, \beta) &= \mathcal{L}(\lambda_0, \beta_0) + \frac{1}{2}\left[\left.\frac{\partial^2 \mathcal{L}}{\partial \lambda^2}\right|_{\lambda_0, \beta_0}(\lambda - \lambda_0)^2 + \left.\frac{\partial^2 \mathcal{L}}{\partial \beta^2}\right|_{\lambda_0, \beta_0}(\beta - \beta_0)^2\right] \quad (5.23) \\
&+ \left.\frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \beta}\right|_{\lambda_0, \beta_0}(\lambda - \lambda_0)(\beta - \beta_0) + \ldots \quad\quad\quad\quad\quad (5.24)
\end{aligned}
$$

This second order term can be expressed in the form of a matrix,

$$
\begin{bmatrix} \lambda - \lambda_0 & \beta - \beta_0 \end{bmatrix}
\begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \lambda^2} & \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \beta} \\ \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \beta} & \frac{\partial^2 \mathcal{L}}{\partial \beta^2} \end{bmatrix}
\begin{bmatrix} \lambda - \lambda_0 \\ \beta - \beta_0 \end{bmatrix}, \quad\quad (5.25)
$$

where the covariance matrix is given as

$$
\begin{bmatrix} \sigma_\lambda^2 & \sigma_{\lambda\beta}^2 \\ \sigma_{\lambda\beta}^2 & \sigma_\beta^2 \end{bmatrix} = -\begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \lambda^2} & \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \beta} \\ \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \beta} & \frac{\partial^2 \mathcal{L}}{\partial \beta^2} \end{bmatrix}^{-1}. \quad\quad (5.26)
$$

The matrix of partial derivatives is known as the Fisher Information Matrix, described by R. A. Fisher in [53]. In order to be used accurately this entity must be calculated with the data being taken into account, however here we are interested not in finding exact results from this method but only in using it to find an appropriate scale for our sampler. By following a procedure of approximation, it is possible to use this to estimate the covariance matrix at any point in the parameter space *if a signal were present at that point*. This is an appropriate quantity to calculate, since we are concerned with finding the size of mode when

there is a signal present, in order to set the right scale.

Approximation of Fisher Matrix

The terms in the Fisher matrix were calculated in an approximate regime as follows. First the log likelihood function is written, presupposing that the data is composed of a signal with normally distributed noise ϵ $N(0, \sigma)$,

$$d(t_j; bfa_0) = s(t_j; \mathbf{a}_0) + \epsilon_j. \tag{5.27}$$

This gives

$$\mathcal{L} \;=\; -N \log\left(\sqrt{2\pi}\sigma\right) + \sum_{j=1}^{N} -\frac{[d(t_j; \mathbf{a}_0) - s(t_j; \mathbf{a})]^2}{2\sigma^2} \tag{5.28}$$

$$=\; E - \frac{1}{2\sigma^2} \sum_{j=1}^{N} \left([s(t_j; \mathbf{a})]^2 - 2s(t_j; \mathbf{a})d_j + d_j^2\right) \tag{5.29}$$

$$=\; E - \frac{1}{2\sigma^2} \sum_{j=1}^{N} \left([s(t_j; \mathbf{a})]^2 - 2hd_j \cos\left[\omega(t_j - \Phi(t_j; \lambda, \beta)) - \phi_0\right] + d_j^2\right) \tag{5.30}$$

Where E is a constant which will be lost through differentiation, as will the d_j^2 term. $[s(t_j; \mathbf{a})]^2$ becomes h^2, minus a sinusoidal term of order h^2 with mean zero. The h^2 disappears on differentiation, and the sinusoidal terms is approximated as summing to zero over the time period and not contributing greatly to the rest of the calculation. The relevant terms to be derived from this are $\partial^2\mathcal{L}/\partial\lambda^2$, $\partial^2\mathcal{L}/\partial\beta^2$ and $\partial^2\mathcal{L}/\partial\beta\partial\lambda$. The first derivatives under the stated assumptions, and where

$f_j = \frac{R}{c} \cos \beta \cos (\lambda - \Omega t_j)$, are found to be [54]

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \frac{h\omega}{\sigma^2} \sum_{j=1}^{N} \left\{ d_j \frac{R}{c} \cos \beta \sin (\lambda - \Omega t_j) \sin [\omega (t_j - f_j) - \phi_0] \right\} \quad (5.31)$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = \frac{h\omega}{\sigma^2} \sum_{j=1}^{N} \left\{ d_j \frac{R}{c} \sin \beta \cos (\lambda - \Omega t_j) \sin [\omega (t_j - f_j) - \phi_0] \right\}. \quad (5.32)$$

The second derivatives are

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda^2} = -\frac{h\omega^2}{\sigma^2} \sum_{j=1}^{N} \left\{ d_j \left[\frac{R}{c} \cos \beta \sin (\lambda - \Omega t_j) \right]^2 \cos [\omega(t_j - f_j) - \phi_0] \right\}$$
$$+ \frac{h\omega}{\sigma^2} \sum_{j=1}^{N} \left\{ d_j \frac{R}{c} \cos \beta \cos (\lambda - \Omega t_j) \sin [\omega(t_j - f_j) - \phi_0] \right\} \quad (5.33)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta^2} = -\frac{h\omega^2}{\sigma^2} \sum_{j=1}^{N} \left\{ d_j \left[\frac{R}{c} \sin \beta \cos (\lambda - \Omega t_j) \right]^2 \cos [\omega(t_j - f_j) - \phi_0] \right\}$$
$$+ \frac{h\omega}{\sigma^2} \sum_{j=1}^{N} \left\{ d_j \frac{R}{c} \cos \beta \cos (\lambda - \Omega t_j) \sin [\omega(t_j - f_j) - \phi_0] \right\} \quad (5.34)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \lambda} = -\frac{h\omega^2}{\sigma^2} \sum_{j=1}^{N} \left\{ \frac{d_j}{4} \left[\frac{R}{c} \right]^2 \sin 2\beta \cos 2 (\lambda - \Omega t_j) \cos [\omega(t_j - f_j) - \phi_0] \right\}$$
$$+ \sum_{j=1}^{N} \frac{h\omega}{\sigma^2} d_j \frac{R}{c} \sin \beta \sin(\lambda - \Omega t_j) \sin [\omega(t_k - f_j) - \phi_0] \quad (5.35)$$

If we recall that at a position of maximum likelihood, the first derivatives must be equal to zero, we can eliminate the second line terms in equations 5.33, 5.34 and 5.35. Since the terms $\cos \beta$ and $\sin \beta$ are independent of j they can be factorised

out of the sum, and as they are both non-zero everywhere [1] this implies that the sum itself is equal to zero. As terms of this sort appear in the second derivatives they may be removed there.

Substituting equation 5.27 into the three second derivatives, and continuing the assumption that we at the true value of the parameters, i.e. $\mathbf{a} = \mathbf{a}_0$, we find

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \lambda^2} &= -\frac{h^2 \omega^2}{\sigma^2} \frac{R^2}{c^2} \cos^2 \beta \\
&\times \sum_{j=1}^{N} \sin^2 (\lambda - \Omega t_j) \cos [\omega(t_j - f_j) - \phi_0] \{\cos [\omega(t_j - f_j) - \phi_0] + \epsilon_j\} \\
\frac{\partial^2 \mathcal{L}}{\partial \beta^2} &= -\frac{h^2 \omega^2}{\sigma^2} \frac{R^2}{c^2} \sin^2 \beta \\
&\times \sum_{j=1}^{N} \cos^2 (\lambda - \Omega t_j) \cos [\omega(t_j - f_j) - \phi_0] \{\cos [\omega(t_j - f_j) - \phi_0] + \epsilon_j\} \\
\frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \beta} &= -\frac{h^2 \omega^2}{4\sigma^2} \frac{R^2}{c^2} \sin 2\beta \\
&\times \sum_{j=1}^{N} \sin 2 (\lambda - \Omega t_j) \cos [\omega(t_j - f_j) - \phi_0] \{\cos [\omega(t_j - f_j) - \phi_0] + \epsilon_j\}
\end{aligned}
$$

Finally, it is noted that any terms involving $\sum_j x_j \epsilon_j$ where x_j is independent of ϵ_j will tend to zero in the limit of large j, as these ϵ_j are drawn from a distribution with mean zero. This approximation removes the reliance on the noise term in the calculation, and it becomes a sum of calculable values only. The evaluation of this is accelerated by approximating the sum as in integral and performing it analytically, removing the need to iterate over the N terms. We arrive at the expressions,

---

[1] Except the three points $\beta \in \{-\frac{\pi}{2}, 0, \frac{\pi}{2}\}$, where the approximation breaks down anyway, giving a singular covariance matrix

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda^2} \approx -\frac{h^2 \omega^2}{4 \Delta T \sigma^2} \frac{R^2}{c^2} \cos^2 \beta$$

$$\times \left([T_{\text{end}} - T_0] + \frac{1}{2\Omega} \left\{ \sin \left[2\lambda - 2\Omega T_{\text{end}} \right] - \sin \left[2\lambda - 2\Omega T_0 \right] \right\} \right) \quad (5.36)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta^2} \approx -\frac{h^2 \omega^2}{4 \Delta T \sigma^2} \frac{R^2}{c^2} \sin^2 \beta$$

$$\times \left(2 \left[T_{\text{end}} - T_0 \right] - \frac{1}{\Omega} \left\{ \sin \left[2\lambda - 2\Omega T_{\text{end}} \right] - \sin \left[2\lambda - 2\Omega T_0 \right] \right\} \right) \quad (5.37)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \beta} \approx -\frac{h^2 \omega^2}{4 \Delta T \sigma^2} \frac{R^2}{c^2} \frac{4}{\Omega} \sin 2\beta \left(\cos \left[2\lambda - 2\Omega T_{\text{end}} \right] - \cos \left[2\lambda - 2\Omega T_0 \right] \right), (5.38)$$

where ΔT is the sampling interval of the data.

These expressions are used with relation 5.26 to find an approximation to the covariance matrix, itself a first order approximate description of the likelihood at any given point. Due to the neglect of the noise in the approximation, the covariance matrices attained this way are slightly smaller than their true values, which is preferable to being slightly too large. The information on the covariance between the parameters does appear successfully in the approximation, so the covariance matrix can still increase the mixing speed of the chain, while preserving the orientation of the mode.

Proposal distributions computed this way have proved effective in increasing the acceptance rate of the proposed steps in sky position from nearly 0% to a level of approximately 10%, which is comparable to the acceptance ratio for steps in the other parameters.

The uniform random distribution is kept, but used only as a first stage proposal function to try large steps, with the new proposal forming the timid stage in the delayed rejection algorithm, similar to the method outlined in section 2.5.2.

Since the proposal distribution varies greatly from location to location in parameter space, the simplification of having equal probabilities of making the forward and reverse move no longer apply. To maintain the principle of detailed balance, the ratio $q_2(\lambda_k, \beta_k | \lambda_{k+1}, \beta_{k+1}) / q_2(\lambda_{k+1}, \beta_{k+1} | \lambda_k, \beta_k)$ must be included in the second stage acceptance probability.

Since the computation of the covariance matrix involves taking the inverse of the Fisher matrix, it is necessary that this matrix is positive definite, i.e. that its determinant be positive. This fails to be true at the points at which $\beta = \{-\frac{\pi}{2}, 0, \frac{\pi}{2}\}$, the two poles and the equator. At these locations, the approximation breaks down and the calculation cannot be performed as intended. This case is tested for in the execution of the code however, and should a matrix inversion problem occur the uniform sampler is used as an alternative.

At positions close to these points also, the covariance matrix can become very large in one direction, reflecting the unresolvability in that parameter. When this occurs, proposals are made outside the range of the β and λ parameters, but these are mapped back into the correct prior range.

5.5 Results of New Search Code

In figure 5.4, a comparison is made between the explicitly calculated posterior probability distribution function and the error ellipse drawn using the covariance matrix as approximated. The scale and orientation of the calculated proposal distribution, while not matching the true distribution perfectly, are sufficient to be useful for the purpose of achieving higher acceptance ratios without incurring significant computational overhead. Such inaccuracies are inevitable given the

level of computational shortcuts which are taken in the approach. There is also the issue that weak signals with short observation times do not in general follow a Gaussian distribution, the higher derivatives in the Taylor expansion becoming more important. However, for most signals the approximation does at least provide an indication of the covariance between the parameters at a particular point in the parameter space, which is an important factor in the mixing of the chain. To produce a more accurate estimation of the mode would require that the data itself be summed over at each proposal, which would greatly reduce the speed of calculation of the proposal densities and therefore the efficiency of the chain. In addition, while the chain is exploring the parameter space, it will not necessarily be located in the peak mode, so the data itself may not produce the required estimate of a mode at the current point if there is a strong signal located elsewhere in parameter space.

5.5.1 Results from MCMC Runs to Estimate Sky Position

The object of this example is to demonstrate the ability of the code to determine the location of the source on the sky. Due to constraints imposed by the computation time necessary as the length of the dataset increases, an observation time of 30 days was used here, with sampling interval 1 minute and variances $\sigma = 1.0$. A signal was injected with parameters $h = 0.6$, $f = 5\,\mathrm{mHz}$, $\phi_0 = \frac{\pi}{2}$, $\lambda = 2.0$ and $\beta = -0.5$. The chain was allowed to run for 1,000,000 iterations, with a thinning factor of 500 and a burn-in length of 100,000 iterations.

The results of the estimation of sky position, shown in figure 5.7 and 5.8, indicate that the Markov Chain has located the mode produced by the original injected
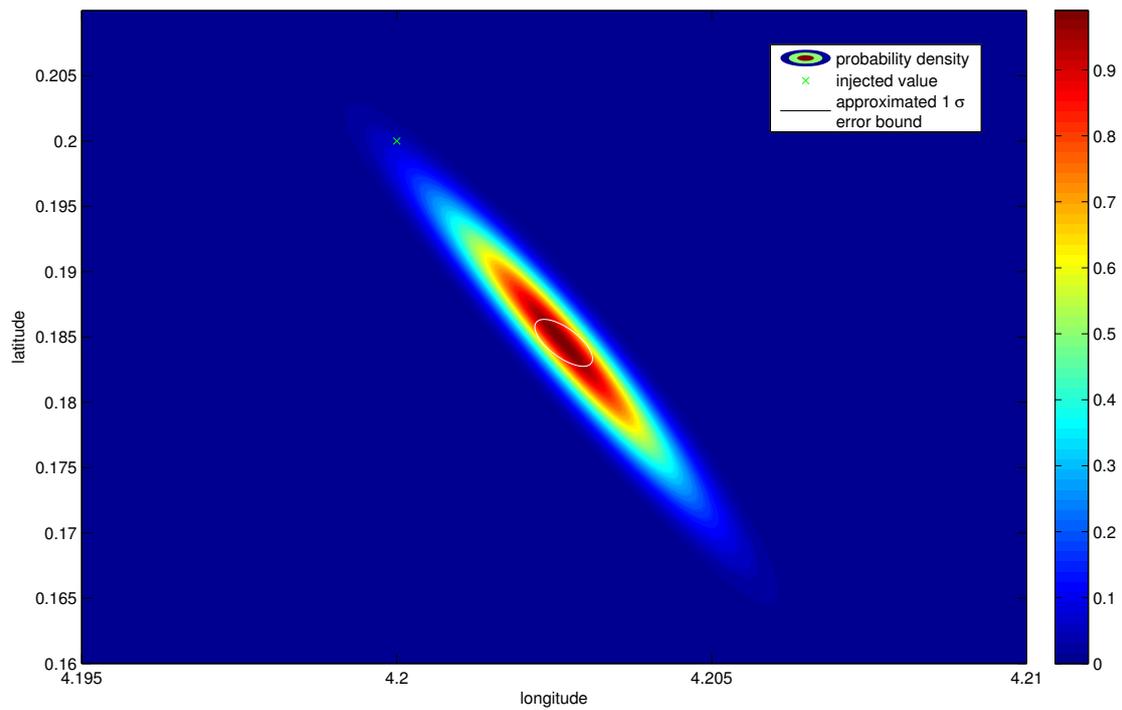
Figure 5.4:   The posterior PDF on sky position for a signal injected into 30 days of white noise $\sigma^2 = 1$ with amplitude $h = 1.0$, $\phi_0 = 0$ and frequency $f = 6\,\mathrm{mHz}$, these parameters being fixed at their true values for the grid-based calculation of the density on sky position. The proposal distribution is shown as a white error ellipse centred on the maximum of the true mode. The approximated ellipse is smaller than the true mode, however this is a desirable feature when the chain has converged and needs to sample from a particular mode.

signal at the initial phase. However, it has also explored the region in parameter space surrounding it, where the maxima in the sky position distribution are shifted under a change of initial signal phase ϕ_0. This indicates a degeneracy in these three parameters which spread the probability distribution out and make it more difficult to localise a signal with a short observation time with this model. The reason for this is clear in equation 5.17, where a change in phase can compensate for a shift in sky position if the observation time is much less than one orbit. Therefore the broad and uneven distribution in phase in figure 5.6 and in the sky positions, which can be seen in figures 5.7 and 5.8.

This problem is alleviated in a full description of the LISA response function, where both the amplitude and frequency modulation depend on sky location, and therefore information about these parameters can be inferred from both sources, where the amplitude modulation is independent of initial phase. Additionally, with extended observation times approaching 1 year, when the modulation of the signal frequency completes a full cycle, its effect cannot be emulated by a shift in initial phase, and the parameters become separable. Unfortunately, a MCMC run over such a lengthy dataset was not possible with the existing code, due to the computation times required to calculate the likelihood. A solution to this problem may be found with the use of parallel computation, as discussed in 5.6.

Despite the effects described above, the absolute amplitude h (figure 5.5) and the frequency f are reproduced very accurately. These can be characterised with the median values and 1σ error bars as $h = 0.593 \pm 0.010$ and $f = 5.00007 \pm 2.3 \times 10^{-5}\,\mathrm{mHz}$.
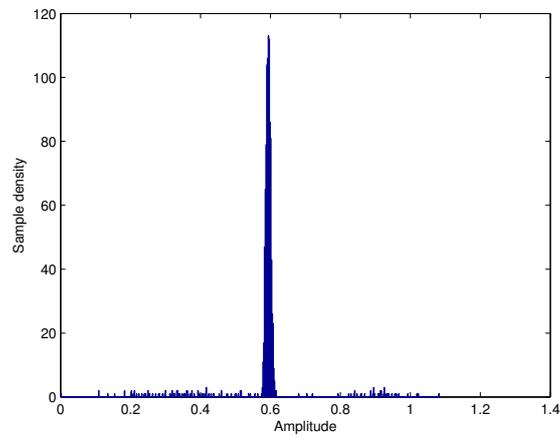
Figure 5.5: The recovered amplitude of the signal has converged well on the value injected at $h = 0.6$.
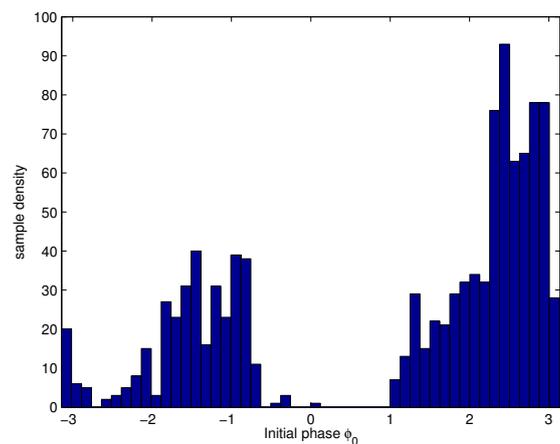
Figure 5.6: The recovered distribution of phase is not accurately estimated, with a broad range of values allowed. This uncertainty in phase allows the sky position to vary outwidth the mode depicted in 5.8, which is computed at the predetermined injected phase. Under a full π shift in ϕ_0, the peaks and troughs of the distribution are exchanged, leading to correlation between the ϕ_0, λ and β parameters. This is shown in figure 5.7.
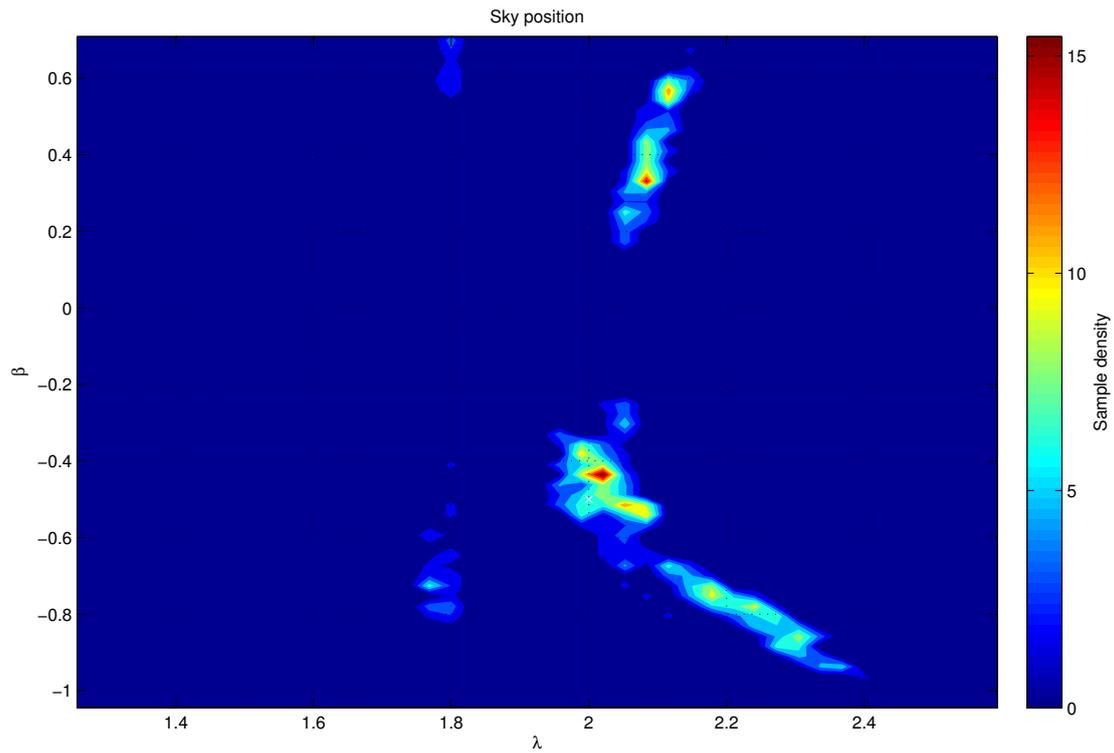
Figure 5.7: The marginal distribution of the chain on the sky position parameters. The chain deviates from the global maximum at the injected parameters $\lambda = 2.0$, $\beta = -0.5$ as the phase changes, however the bulk of the probability is localised in the mode around the maximum shown in figure 5.8.
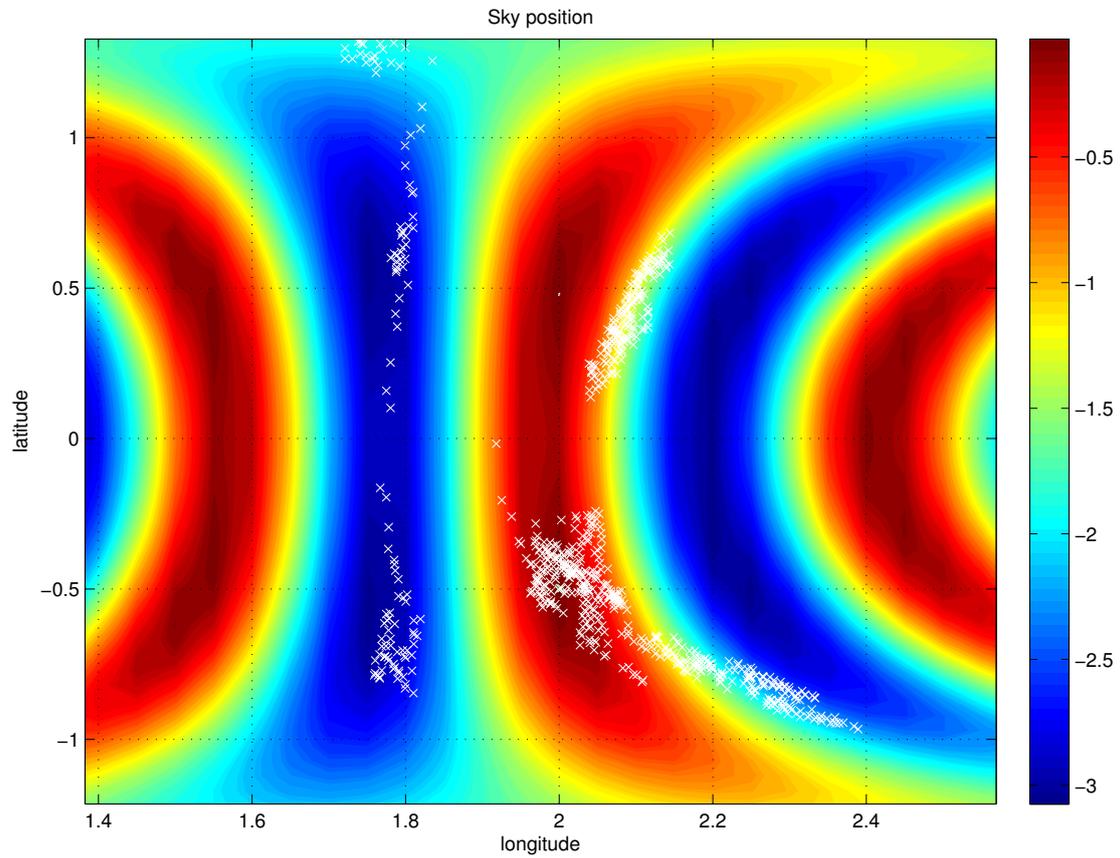
Figure 5.8: The log likelihood posterior PDF over sky position, maximised over all other parameters. The injected value at longitude 2.0, latitude -0.5 is marked with the symbol ×. The samples from the Markov chain are shown as white crosses, which indicate that the chain has explored both the true maximum mode and the alternative adjacent mode produce when the initial phase is shifted by π.

5.5.2 Testing Sensitivity of the Algorithm

As opposed to the pulsar MCMC search, the LISA binary search has the ability to use a model with no signal at all. This feature removes the problem of a signal being matched to the slight sinusoidal features in the noise when there is no strong signal present. In this test I shall examine the reversible jump behaviour of the algorithm as the signal to noise ratio approaches zero, and the probability of a signal being present decreases accordingly.

A series of signals were injected into white Gaussian noise files, $\sigma = 1.0$ and duration 1000 minutes, with the usual sampling interval of 1 minute. Each MCMC run was started at a random point in parameter space, and allowed to evolve for $10\,000\,000$ iterations with a burn-in time of $100\,000$ samples and a thinning factor of 500. The greater number of iterations here than in the pulsar search is to allow the chain to explore the multidimensional parameter space which spans different models.

Since transitions can be made between models with differing signal numbers, there is a more gradual transition between a detectable signal and a non-detectable one than in the case of the pulsar code in 3.1.2. However, in this case there is no equivalent of the $\cos \iota$ parameter to affect the signal to noise ratio, as the effects of the source orientation are not included in the model. The variable number of signals in this code allows us to set a direct Bayesian upper limit of 95% confidence without the need to resort to large numbers of Monte Carlo injections. While in practice the signals in LISA are expected to have a high signal to noise ratio, it is interesting to perform this analysis to test the sensitivity of the algorithm.

In figure 5.9 is shown the probability assigned to each model as a function
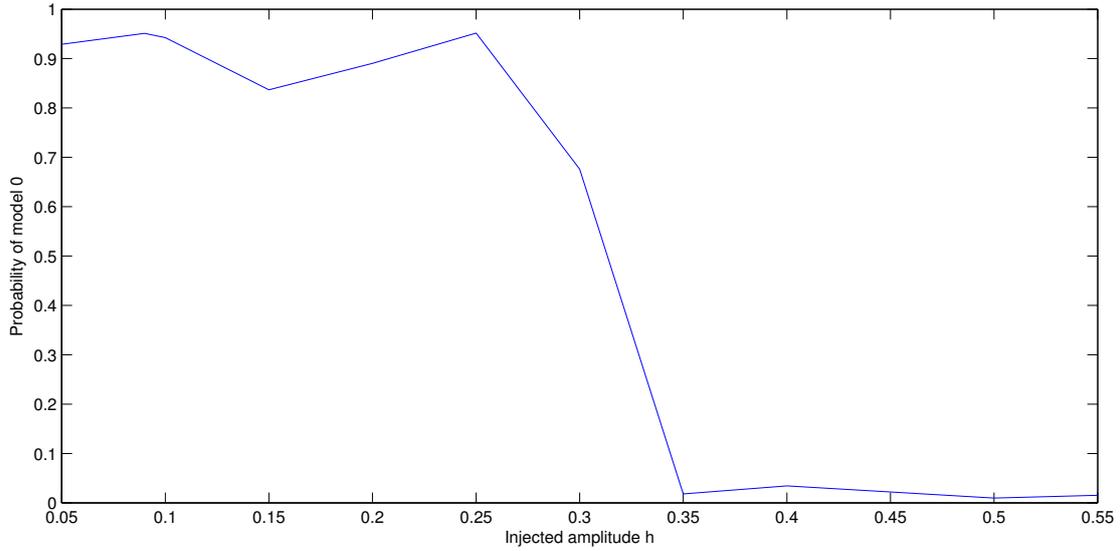
Figure 5.9: The decreasing probability of the null hypothesis $\mathcal{M}_0$ with no signal modeled, as a function of increasing injected signal amplitude. The point at which the model becomes only 5% probable is $h = 0.346$, corresponding to a signal to noise ratio of 10.9.

of injection amplitude h. When the injection reaches a value of $h = 0.346$ the probability of the model with no signals has reduced to 5%, corresponding to the upper limit in section 3.1.2. For comparison, this amplitude is equivalent to a signal to noise ratio of 10.9. This indicates that the more advanced LISA algorithm is greatly superior in detecting low-level signals. Due to its ability to evaluate the null hypothesis, it is not forced to fit a signal model to the data even when there is no signal present as the pulsar algorithm is. This allows the Markov chain to avoid getting stuck in the local maxima of probability, a feature that would provide significant benefit in the case of assigning upper limits, rather than estimating parameters.

5.5.3 Estimating the Number of Sources

Here we will demonstrate the ability of the algorithm to infer the number of sources present in the dataset, given sufficient signal power, and estimate the levels of noise accordingly. This test depends on the correct behaviour of the transdimensional jumps between the set of models $\mathcal{M}_m$ and requires a lengthier burn-in stage to bring the chain into the region of the most probable models. The density of samples in each model $\mathcal{M}_m$ is then proportional to the probability of each of these models.

Test data

The dataset used in this example contained 50 test signals injected into 4000 minutes of Gaussian white noise of variance $\sigma^2 = 1$. The differing amplitudes of these 50 signals were chosen to explore the limits at which the algorithm no longer assigns any probability to the model and instead allocates its power to the noise. To this end the fifty amplitudes were evenly spaced from $h = 0.0$ to $h = 1.0$. The signals were also distributed evenly in frequency to allow them to be distinguished easily, and here the range was between $f = 1\,\mathrm{mHz}$ and $f = 7.333333\,\mathrm{mHz}$. The dataset used is represented in the frequency domain in figure 5.10.

Results

The sampler was run for a total of $10\,000\,000$ iterations, the first $1\,000\,000$ of which were a burn-in stage. The thinning factor was 500, leaving $18\,000$ samples from which the PDF was estimated.
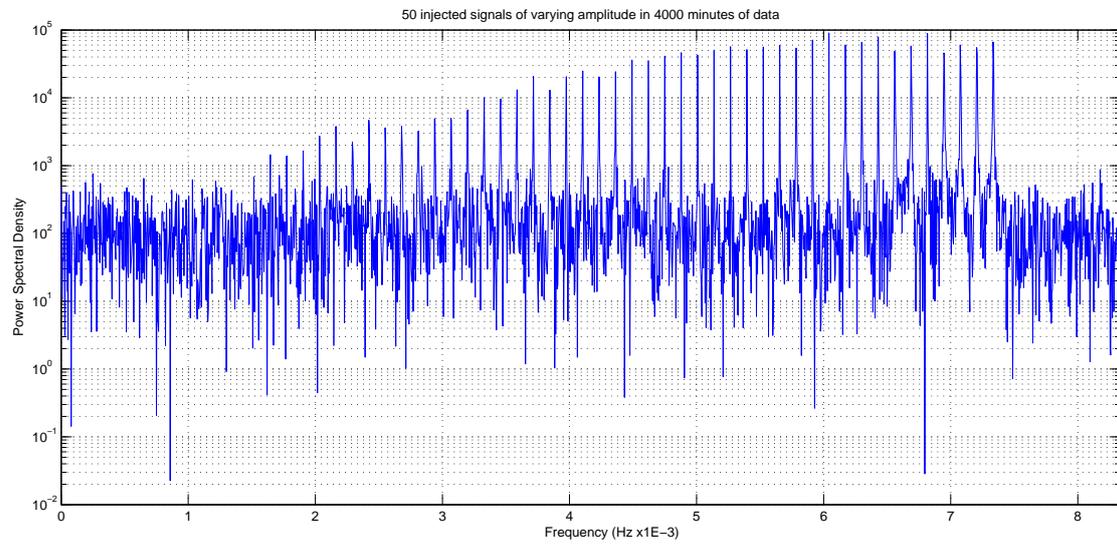
Figure 5.10: The 50 injected signals, displayed in the frequency domain.
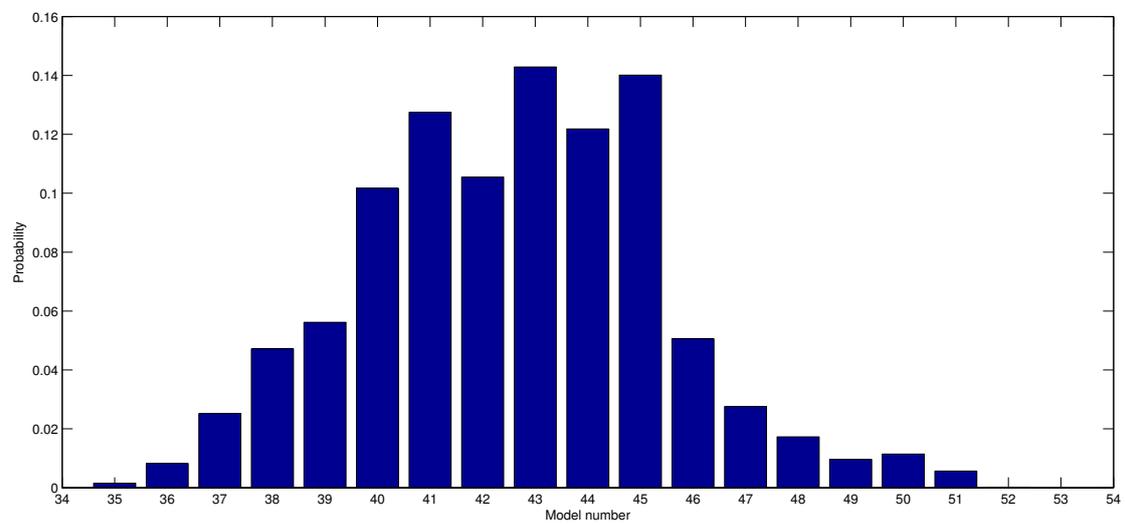
Figure 5.11: The probabilities of each of the models $\mathcal{M}_m$ as estimated from the MCMC sampler.

Figure 5.11 shows the marginalised probability distribution over model number $\mathcal{M}_m$, and figure 5.12 shows the estimate of the noise σ_m for each model $\mathcal{M}_m$. The noise estimate varies with the number of signals, when higher numbers of signals are estimated there is less power allocated to the noise estimate, and this is therefore lower. Conversely, at lower model numbers there is a higher estimate of the noise level accordingly. As the lowest amplitude signals were not detected by the algorithm, the estimated number of signals include sporadic extra signals which were not present in the injection but contribute to the posterior by increasing m, as they can be accepted briefly and eliminated again. There is also the possibility of two large amplitude signals with differing phase or sky position interfering to generate a waveform which provides a close fit to the data. This can occur when a single signal undergoes a split transition, and the two resultant signals closely reproduce the original waveform.

5.6   Comments and Further Work

The method outlined above demonstrates the first step in developing a possible approach to LISA data analysis which has a real chance of tackling the source confusion problem. The successful use of reversible jump MCMC to move between models with different numbers of parameters offers a way to algorithmically and automatically parametrise those sources which are distinguishable in the data. The remaining gravitational wave power from unresolvable sources is folded into the estimate of the noise level, which is itself inferred from the data. Bayesian inference provides a robust approach to finding the probability density of models
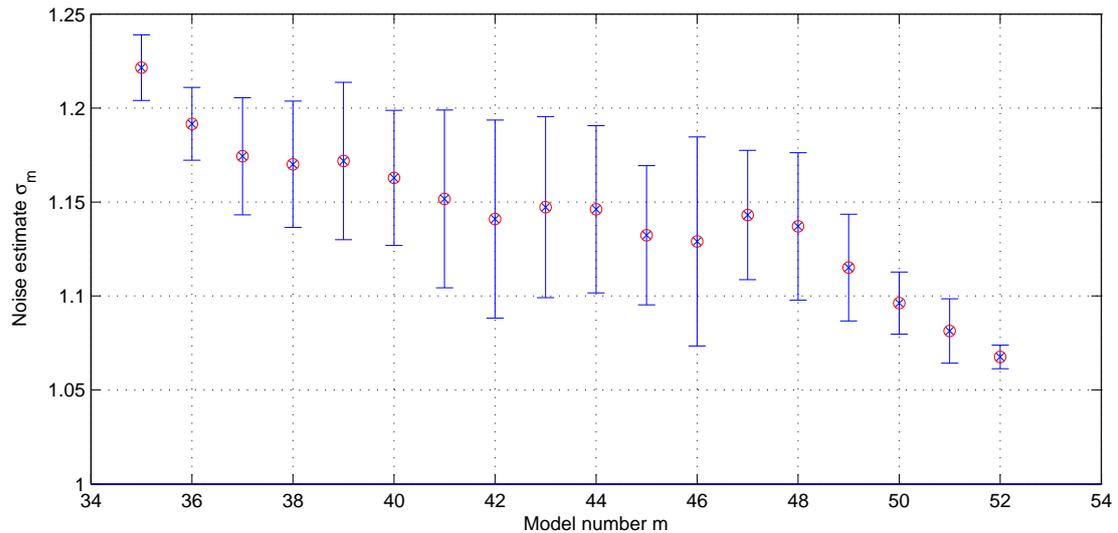
Figure 5.12: The estimate of noise level σ, displayed as a function of model number alone, with the standard deviation of the estimate indicated by the error bars. The simulated noise variance was $\sigma = 1$. At larger values of m the noise estimate is closer to the injected noise level, as more power is accounted for by the signal models. Note that the estimates for model number 52 were calculated from only 3 samples with low probabilities. With some signals not providing enough power for their presence to be detected, the models with high numbers of signals may include those where the model number is overestimated by the inclusion of small amplitude signals which individually may only last for a few samples as they are of low probability, and of pairs of signals which interfere to produce a low amplitude signal, possibly resulting from a *split* transition.

with differing numbers of sources, a key requirement in the situation where the number of sources cannot be determined from any external observations.

The simplistic nature of the model, however, is a limiting factor in the usefulness of this particular implementation. The shape of the multimodal distribution on sky position arises from the symmetry between the north and south hemispheres, and although the approximation I have developed is useful for exploring a mode it is not good at moving between the concentric rings of probability as these lie in the direction perpendicular to the alignment of the mode.

It is clear from the above examples that a more accurate modelling of the LISA detector and response function is required to improve the performance of the algorithm in resolving signals on the sky. This would break the degeneracies and better localise probability on the true origin of the signal. Although still in the early stages of development at the time of writing, the ongoing work in this area by myself, and recently similar approaches by others in the field promise to provide an important part of LISA data analysis. The basic details of the algorithm such as its use of Markov Chain Monte Carlo, reversible jumps and delayed rejection has been taken up by others in the gravitational wave data analysis community, and similar approaches have been used in the analysis of the first round of the Mock LISA Data Challenges (MLDC) [14].

A hierarchical approach which involves an exhaustive search, followed by a MCMC sampler for parameter estimation has been developed for these challenges in collaboration with Alexander Ströer *et al* [55] and is part of an ongoing participation in the MLDC. This approach is based in the frequency domain, allowing the use of the Fast Fourier Transform to improve performance and allow analysis of the full datasets provided as part of the MLDC, which have durations greater

than 1 year.

An independent line of work on the same problem has also been pursued by Neil Cornish and Jeff Crowder, which tackles the problem using an MCMC algorithm but does not freely explore the number of models in the same way, details of which can be found in [56].

The object of this work was to demonstrate a proof of concept implementation of Bayesian model selection in a reduced version of the LISA source confusion problem. With the increasing power of computers and the continued development of faster algorithms I believe that in the future the full problem will become tractable through means such as this. Statistical inference will also form an important part of the approaches to other data analysis challenges with LISA, as it offers the best way of extracting maximum scientific information from the observations.

Chapter 6

Concluding remarks

In recent years, the field of astronomy has increasingly made use of probabilistic inference in the analysis of observations, particularly in the area of cosmology, where competing models are tested against the limited available data. As the study of gravitational waves progresses to becoming another branch of astronomy, their extremely weak nature will necessitate the use of the most sensitive techniques to gain the maximum scientific benefit from the noisy data. The development of data analysis algorithms based on Bayesian inference is therefore important for the advancement of the field, and the work which has been presented here is a small part of that development.

The full extraction of information about a model from observations can require the exploration of large parameter spaces as the models become more complex. I have found that Markov chain Monte Carlo can be a powerful technique for examining such probability distributions that have a high dimensionality, where exhaustive methods are not possible even on modern computing hardware. It should be emphasised however that while the basic technique of MCMC as outlined

in section 2.4.1 is conceptually simple, its generality means that optimising it for a particular task can be a difficult problem in its own right. As we have seen in the approach to pulsar data analysis in chapters 3 and 4, the use of a probabilistic algorithm comes at the cost of sensitivity in comparison to exhaustive methods in problems where the bulk of probability is concentrated in a small corner of parameter space. Given that this corner can occupy less than 1 part in 10^{10} of the total hyper-volume, as in section 3.1, it is perhaps surprising the technique works as well as it does. That is not to say that the potential of MCMC has been fully explored here. Indeed, the broad field of Markov chain Monte Carlo algorithms is still undergoing rapid development, and techniques invented in recent years such as delayed rejection and the reversible jump method of moving between models of variable dimensionality have further increased the power of this approach. In chapter 5 these methods were applied to the problem of LISA source confusion, where it is likely that some form of model selection will be required to allow meaningful inferences to be drawn on the observations. The success in applying these methods to the reduced problem that has been tackled shows the promise that they hold, in addition to the requirement of tuning the algorithm for a particular task as exemplified by the work on adaptive proposal distributions in section 5.4.3. This has been confirmed at the time of writing by the prominence of MCMC methods in the more recent first round of Mock LISA Data Challenges, where the difficulty of the analysis task has forced the use of advanced implementations of the method. It has also found a place as part of a pipeline approach to LISA data analysis, where it can be used to improve the parameter estimates of candidate sources found by an initial stage [55].

The analysis of data from the LIGO H1 interferometer during the S3 science

run has produced the upper limit on gravitational wave radiation from SN1987a in a 4 Hz by 2×10^{-10} Hzs^{-1} window around 935Hz of $h_{95\%} = 7.3 \times 10^{-23}$ in section 4.3. The search and upper limit estimation was performed in less than 87,000 hours of CPU time on a 2 GHz cluster, more than ten times faster than the commonly used $\mathcal{F}$-statistic algorithm, at the cost of decreased sensitivity. The actual search stage took up only 6.6% of this time, the rest being used to set the upper limit. The performance of the MCMC algorithm used in chapter 5 for LISA data analysis suggests that use of a transdimensional jump method could be extremely beneficial in the case of the pulsar search also, offering both an large improvement in sensitivity and the elimination of the time-consuming Monte Carlo injections to find upper limits.

The continued improvements in the efficiency and reliability of MCMC methods make them likely to see increased usage in the field of gravitational wave data analysis as it develops from a search to a study. The extraordinary technical achievement of constructing instruments as sensitive as the LIGO and GEO interferometers has for the first time delivered a real possibility of detecting a gravitational wave. It is my hope that the data analysis community will continue to find innovative ways to meet the challenges posed by these experiments, and by future generations of interferometers both on Earth and in space. When we have achieved the detection of gravitational waves, we will have begun a new way of doing astronomy, allowing the study of the universe to proceed in directions never before possible.

Bibliography

[1] Peter Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711, 1995.

[2] R. Umstätter, N. Christensen, M. Hendry, R. Meyer, V. Simha, J. Veitch, S. Vigeland, and G. Woan. Bayesian modeling of source confusion in LISA data. *Phys. Rev. D*, 72(2):022001–+, July 2005.

[3] Bernard F. Schutz. *A First Course in General Relativity*. Cambridge University Press, 1985.

[4] Misner, Thorne, and Wheeler. *Gravitation*. Freeman, 1973.

[5] S. Hawking and W. Israel. *300 Years of Gravitation*. Cambridge University Press, 1987.

[6] Graham Woan. *The Cambridge Handbook of Physics Formulas*. Cambridge University Press, 2000.

[7] LIGO Scientific Collaboration:B. Abbott. Search for gravitational waves from binary black hole inspirals in LIGO data. *Phys. Rev. D*, 73(6):062001–+, March 2006.

[8] J. W. T. Hessels, S. M. Ransom, I. H. Stairs, P. C. C. Freire, V. M. Kaspi, and F. Camilo. A radio pulsar spinning at 716 hz. *Science*, 311:1901–1904, March 2006.

[9] N. Andersson and K. D. Kokkotas. Towards gravitational wave asteroseismology. *Monthly Notes of the Royal Astronomical Society*, 299:1059–1068, October 1998.

[10] G. Ushomirsky, C. Cutler, and L. Bildsten. Deformations of accreting neutron star crusts and gravitational wave emission. *Monthly Notices of the Royal Astronomical Society*, 319:902–932, December 2000.

[11] J. M. Weisburg and J. H. Taylor. Relativistic binary pulsar B1913+16: Thirty years of observations and analysis. *Binary Radio Pulsars, Aspen Conference Series*, CS-328, 2005.

[12] R. A. Hulse and J. H. Taylor. *Astrophys. J.*, 195:L51, 1975.

[13] M. J. Benacquista, J. DeGoes, and D. Lunder. A simulation of the laser interferometer space antenna data stream from galactic white dwarf binaries. *Classical and Quantum Gravity*, 21:509–+, March 2004.

[14] K. A. Arnaud *et al.* Report on the first round of the mock lisa data challenges. 2007. http://arxiv.org/abs/gr-qc/0701139 *Submitted to Classical and Quantum Gravity.*

[15] K. A. Arnaud *et al.* An overview of the second round of the mock lisa data challenges. January 2007. http://arxiv.org/abs/gr-qc/0701170 *Proceedings of GWDAW 11.*

[16] I. Ciufolini, V. Gorini, U. Moschella, and P. Fré. *Series in High Energy Physics, Cosmology and Gravitation: Gravitational Waves.* Insitute of Physics Publishing, 2001.

[17] M. Maggiore. Gravitational wave experiments and early universe cosmology. *Phys. Rep.*, 331:283–367, 2000.

[18] LIGO Scientific Collaboration. Searching for a Stochastic Background of Gravitational Waves with LIGO. *ArXiv Astrophysics e-prints:0608606*, August 2006.

[19] B. Allen. The stochastic gravity-wave background: sources and detection. *Proceedings of the Les Houches School on Astrophysical Sources of Gravitational Waves*, 1996.

[20] P. Jaranowski, A. Królak, and B. F. Schutz. Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection. *Phys. Rev. D*, 58(6):063001–+, September 1998.

[21] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53:370–418, December 1763.

[22] Harold Jeffreys. *Scientific Inference.* Cambridge University Press, 1931.

[23] E. T. Jaynes. *Probability Theory: the Logic of Science.* Cambridge University Press, 2003.

[24] D. S. Sivia. *Data Analysis: A Bayesian Tutorial.* Oxford University Press, 1996.

[25] M. Zimmermann and E. Szedenits, Jr. Gravitational waves from rotating and precessing rigid bodies - Simple models and applications to pulsars. *Phys. Rev. D*, 20:351–355, July 1979.

[26] Rejean Dupuis, Graham Woan, and the LIGO scientific collaboration. Limits on the gravitational wave strength from isolated radio pulsars using data from the second science run of ligo. Technical report, LIGO-P040008-00-Z, 2004.

[27] Matthew Pitkin. *Searches for continuous and transient gravitational waves from known neutron stars and their astrophysical implications.* PhD thesis, University of Glasgow, 2005.

[28] Rejean J. Dupuis. *Bayesian Searches for Gravitational Waves from Pulsars.* PhD thesis, University of Glasgow, 2004.

[29] Dani Gamerman. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Chapman and Hall, 1997.

[30] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1090, 1953.

[31] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[32] Phil Gregory. *Bayesian Logical Data Analysis for the Physical Sciences.* Cambridge University Press, 2005.

[33] R. Umstätter, R. Meyer, R. J. Dupuis, J. Veitch, G. Woan, and N. Christensen. Detecting Gravitational Radiation from Neutron Stars using a Six-Parameter Adaptive MCMC Method. In R. Fischer, R. Preuss, and U. V. Toussaint, editors, *American Institute of Physics Conference Series*, pages 336–343, November 2004.

[34] Luke Tierney and Antonietta Mira. Some adaptive monte carlo methods for bayesian inference. *Statistical Medicine*, 18:2507–2515, 1999.

[35] J. Middleditch, J. A. Kristian, W. E. Kunkel, K. M. Hill, R. D. Watson, R. Lucinio, J. N. Imamura, T. Y. Steiman-Cameron, A. Shearer, R. Butler, M. Redfern, and A. C. Danks. Rapid photometry of supernova 1987A: a 2.14 ms pulsar? *New Astronomy*, 5:243–283, August 2000.

[36] Lal home page: http://www.lsc-group.phys.uwm.edu/daswg/projects/lal.html.

[37] Jolien Creighton. Lal sofware documentation: http://www.lsc-group.phys.uwm.edu/daswg/projects/lal/docs/lsd-5.1.pdf.

[38] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472, Nov. 1992.

[39] G. Larry Bretthorst. *Bayesian Spectrum Analysis and Parameter Estimation.* Springer-Verlang, 1988.

[40] E.T. Jaynes. Bayesian spectrum and chirp analysis. *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*, 1987.

[41] John Veitch, Richard Umstaetter, Renate Meyer, Nelson Christensen, and Graham Woan. A time-domain mcmc search and upper limit technique for gravitational waves of uncertain frequency from a targeted neutron star. *Classical and Quantum Gravity*, 22:S995, 2005.

[42] Condor home page: http://www.cs.wisc.edu/condor/.

[43] Matthew Pitkin. Private communication, 2006.

[44] W. D. Arnett, J. N. Bahcall, R. P. Kirshner, and S. E Woosley. Supernova 1987a. *Ann. Rev. Astron. Astrophys.*, 27:629–700, 1989.

[45] Middleditch J., J. A. Kristian, W. E. Kunkel, Hill K. M., and R. D. Watson. A 2.14 ms candidate optical pulsar in sn1987a. *ArXiv Astrophysics e-prints:0010044*, October 2000.

[46] K. Soida, M. Ando, N. Kanda, H. Tagoshi, D. Tatsumi, K. Tsubono, and the TAMA Collaboration. Search for continuous gravitational waves from the SN1987A remnant using TAMA300 data. *Classical and Quantum Gravity*, 20:645–+, September 2003.

[47] Chris Messenger. Private communication, 2007.

[48] Christopher Messenger. *Gravitational Wave Data Analysis, Searching For Gravitational Waves from Low Mass X-Ray Binaries.* PhD thesis, University of Birmingham, 2006.

[49] Karsten Danzmann and Albrecht Ruïger. Lisa technology - concept, status, prospects. *Classical and Quantum Gravity*, 20:S1–S9, 2003.

[50] R. Umstätter, N. Christensen, M. Hendry, R. Meyer, V. Simha, J. Veitch, S. Vigeland, and G. Woan. LISA source confusion: identification and characterization of signals. *Classical and Quantum Gravity*, 22:901–+, September 2005.

[51] T. A. Prince, M. Tinto, S. Larson, and J. W. Armstrong. Lisa optimal sensitivity. *Phys. Rev. D*, 66:122002, 2002.

[52] J. W. Armstrong, F. B. Estabrook, and M. Tinto. Time-Delay Interferometry for Space-based Gravitational Wave Searches. *Ap. J.*, 527:814–826, December 1999.

[53] R. A. Fisher. *J. Roy. Stat. Soc.*, 98:39, 1935.

[54] Luis Theodoro. Private communication, 2005.

[55] Alexander Stroeer, John Veitch, Christian Roever, Ed Bloomer, James Clark, Nelson Christensen, Martin Hendry, Chris Messenger, Renate Meyer, Matthew Pitkin, Jennifer Toher, Richard Umstaetter, Alberto Vecchio, and Graham Woan. Inference on white dwarf binary systems using the first round mock lisa data challenges data sets. 2007. http://arxiv.org/abs/0704.0048 *Submitted to Classical and Quantum Gravity*.

[56] J. Crowder and N. J. Cornish. Extracting galactic binary signals from the first round of Mock LISA Data Challenges. *Classical and Quantum Gravity*, 24:575–+, October 2007.