A Metropolis–Hastings routine for estimating parameters from compact binary inspiral events with laser interferometric gravitational radiation data

Nelson Christensen[1], **Renate Meyer**[2] **and Adam Libson**[1]

[1] Physics and Astronomy, Carleton College, Northfield, MN 55057, USA
[2] Department of Statistics, The University of Auckland, Auckland, New Zealand

E-mail: nchriste@carleton.edu, meyer@stat.auckland.ac.nz and libsona@carleton.edu

Abstract
Presented here are the results of a Metropolis–Hastings Markov chain Monte Carlo routine applied to the problem of determining parameters for coalescing binary systems observed with laser interferometric detectors. The Metropolis–Hastings routine is described in detail, and examples show that signals may be detected and analysed from within noisy data. Using the Bayesian framework of statistical inference, posterior distributions for the parameters of the binary system are derived using our routine.

PACS number: 04.80.Nn

1. Introduction

This is an exciting time to study gravitation, astrophysics and cosmology. Through challenging cosmic microwave background (CMB) and supernovae observations cosmology has been turned on its head. Gravitational radiation astronomy should be the next contributor to this revolution in astrophysics and cosmology. The attempt to detect gravitational radiation is a difficult and exciting area of current research. LIGO has been built and is operating [1, 2]. Around the globe a worldwide network of detectors is coming online; TAMA in Japan [3] and GEO in Germany [4] are operating alongside LIGO in the quest for gravitational radiation detection, while VIRGO in Italy [5, 6] will soon also be online. The existence of gravitational radiation was predicted by Albert Einstein [7], and confirmed through the binary neutron star observations of Taylor and Weisberg [8]. The direct detection of gravitational radiation will be a tremendous accomplishment for physics, but it will almost certainly also cause an upheaval in our knowledge of the universe and how it works. The observations of expected sources should be dramatic: supernovae, pulsars, the inspiral of binary neutron stars followed by black-hole formation, or even the stochastic background from the Big Bang.

Coalescing binaries containing neutron stars or black holes are expected to be the cleanest and most promising source of detectable radiation [9]. The detection of numerous coalescing binary events may provide information of cosmological significance [10], such as an independent measurement of the Hubble constant [11–13]. Observation of the time of tidal disruption of neutron star binary systems may permit a determination of the neutron star radii and information on the equation of state [14–16]. The characteristics of radiation in the post-Newtonian regime will provide insight into highly nonlinear general relativistic effects, such as the observation of the formation of a Kerr black hole as the binary system decays [12, 17, 18].

A Bayesian approach to parameter estimation is well suited to astrophysical studies [19, 20]. The inevitable difficulty in calculating the multidimensional integrals necessary for the determination of posterior probability distributions has hindered the further development of Bayesian parameter estimation methods in astrophysics despite their well-known advantages over conventional frequentist techniques [20–23]. These impediments have been overcome as the last two decades have seen a revolution in Bayesian computational techniques through the enormous progress made in computer technology and the development of simulation-based Markov chain Monte Carlo (MCMC) methods [24–26]. Previously, we investigated whether a MCMC-based Bayesian approach could prove useful for analysing gravitational radiation data [27]. We used a specific MCMC method, the Gibbs sampler [28], to show that parameter estimation and posterior distribution generation can be accomplished for compact binary inspiral signals detected by laser interferometric antennas [29]. However, it is well known that the Gibbs sampler is inefficient when parameters are highly correlated. Also, our Gibbs sampler demonstration employed an existing and specialized MCMC software called BUGS (Bayesian inference using Gibbs sampling) [30] which cannot be incorporated into the LIGO/LSC [31] algorithm library (LAL) [32]. Thus, in the study presented in this paper we developed and applied a Metropolis–Hastings (MH) [33, 34] routine that is designed to eventually operate within LAL [32]. The MH routine we use for the present binary inspiral gravitational radiation analysis is based on the same rationale as the one we developed for extracting cosmological parameters from cosmic microwave background data [35, 36].

Signal detection, parameter estimation and statistical analysis from interferometer data present difficult and challenging projects. MCMC methods are making significant contributions to difficult data analysis problems in many disciplines, from science to economics [24], where the number of parameters is large. Ultimately the analysis of interferometer data for binary inspiral events will have to account for a large number of parameters. For a single interferometer detection, gravitational radiation researchers are concentrating their efforts on identifying coalescing binary events with just five parameters; the two masses, the effective distance from Earth to the binary system (or equivalently the signal amplitude), the orbital phase and time at coalescence. With a network of detectors there are more parameters to be found, such as source location in the sky, and polarization of the gravitational radiation. Real events will likely be influenced by the interaction of the spin angular momentum of each compact object with the orbital angular momentum [37–40]. MCMC methods may be the best hope for exploring the complicated phase space for binary inspiral gravitational radiation signals when the spins of the masses are included in the detection template. A similar effort has been made in creating MCMC inspiral event code that models the final plunge of the two compact objects towards one another [41–44], and the subsequent quasi-normal mode ringing of the newly formed black hole. A worldwide network of detectors will ultimately be used to observe and record compact binary inspiral events [45–47]. MCMC methods offer a realistic means to extract the large number of parameters from gravitational radiation data from complicated binary inspirals detected by a worldwide network of detectors. In this paper

we demonstrate the potential of the MH algorithm for simulated data from a binary inspiral (neglecting spin and final plunge) detected by a single detector. We are expanding our routine to account for more complex waveforms and multiple detectors; that work is in progress.

This paper is organized as follows: in section 2 we briefly review Bayesian inference. The description of the MCMC simulation technique we use, specifically the Metropolis–Hastings routine, is presented in section 3. In section 4 we present two examples where we use our MCMC approach to identify the parameters which created the signal that is buried in synthesized noise. We conclude with a discussion of our results and the direction of future efforts in section 5.

2. Bayesian inference

Presented here is a review of the Bayesian approach to parameter estimation that we pursued in this study. A good comparison of *frequentist* versus *Bayesian* statistics, as applied to gravitational radiation detection, can be found in the work of Finn [48]. Call the antenna output data $\mathbf{z}(t)$. One starts with the *likelihood* function $p(\mathbf{z}|\boldsymbol{\theta})$, namely the conditional probability density function (PDF) of the observation $\mathbf{z}$, conditional on unobserved parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. The *likelihood* $p(\mathbf{z}|\boldsymbol{\theta})$ is regarded as a function of the parameters $\boldsymbol{\theta}$. The likelihood has been used extensively in gravitational radiation data analysis and parameter estimation research (for example with respect to coalescing binaries) [12, 13, 47, 49–52]. In contrast to the frequentist approach where $\boldsymbol{\theta}$ is regarded as fixed but unknown, the Bayesian approach treats $\boldsymbol{\theta}$ as a random variable with a probability distribution that reflects the researcher's uncertainty about the parameters. Bayesian inference requires the specification of a prior PDF for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$, that quantifies the researcher's pre-experimental uncertainty about $\boldsymbol{\theta}$. It should take all information into account that is known about $\boldsymbol{\theta}$ before observing the data. All information about $\boldsymbol{\theta}$ that stems from the experiment should be contained in the likelihood but cannot be used to specify the prior distribution. Bayesian inference then uses the experimental information through the likelihood to update the prior PDF. Via an application of Bayes' theorem, by conditioning on the known observations, this post-experimental knowledge about $\boldsymbol{\theta}$ is expressed through the *posterior* PDF

$$p(\boldsymbol{\theta}|\mathbf{z}) = \frac{p(\boldsymbol{\theta})\,p(\mathbf{z}|\boldsymbol{\theta})}{m(\mathbf{z})} \propto p(\boldsymbol{\theta})\,p(\mathbf{z}|\boldsymbol{\theta}), \tag{1}$$

where $m(\mathbf{z}) = \int p(\mathbf{z}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$ is the marginal PDF of $\mathbf{z}$ which can be regarded as a normalizing constant as it is independent of $\boldsymbol{\theta}$. The posterior PDF is thus proportional to the product of prior and likelihood.

The standard Bayesian point estimate of a single parameter, say θ_i, is the posterior mean

$$[\theta_i] = \int \theta_i\, p(\theta_i|\mathbf{z})\,\mathrm{d}\theta_i, \tag{2}$$

where

$$p(\theta_i|\mathbf{z}) = \int \cdots \int p(\boldsymbol{\theta}|\mathbf{z})\,\mathrm{d}\theta_1 \cdots \mathrm{d}\theta_{i-1}\,\mathrm{d}\theta_{i+1} \cdots \mathrm{d}\theta_d \tag{3}$$

is the marginal posterior PDF obtained by integrating the joint posterior PDF over all other components of $\boldsymbol{\theta}$, except θ_i. A measure of the uncertainty of this estimate is the posterior standard deviation, or a 95% credibility interval that contains the parameter θ_i with 95% probability, its lower and upper bounds being specified by the 2.5% and 97.5% percentile of $p(\theta_i|\mathbf{z})$, respectively.

3. Posterior computation via MCMC

As seen from equations (2) and (3), the calculation of posterior means requires d-dimensional integration, one of the main issues that has made the application of Bayesian inference so difficult in the past. This hurdle has been overcome by the great advances in simulation-based integration techniques. The basic simulation-based approach to calculating posterior means is to simulate, say K, independent values, $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(K)}$, from $p(\boldsymbol{\theta}|\mathbf{z})$. A consistent estimate of equation (2) is then given by the sample average

$$\hat{\theta}_i = \frac{1}{K} \sum_{k=1}^{K} \theta_i^{(k)}. \tag{4}$$

However, if the dimension d of the parameter space is large, sampling directly from $p(\boldsymbol{\theta}|\mathbf{z})$ is a formidable problem. MCMC methods [24, 27, 29] have been developed to tackle this. One method for constructing a Markov chain is the Gibbs sampler [28], which we employed previously [29]. It reduces the problem of sampling from a d-dimensional PDF $p(\boldsymbol{\theta}|\mathbf{z})$, to simulating in, a cyclic way, from d one-dimensional PDFs, the full conditional PDFs $p(\theta_i|\mathbf{z}, \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d)$.

A more general method is the Metropolis–Hastings algorithm [33, 34]. The Metropolis–Hastings algorithm is among the top of the list of great algorithms of 20th century scientific computing [53], and widely used for simulation in various disciplines such as biology, chemistry and economics. The Gibbs sampler, a special case of the MH algorithm, forms the backbone of the popular statistical software package BUGS [30].

We first describe the original Metropolis algorithm [33] and then its generalization due to Hastings [34], and give an explanation of why this algorithm correctly samples from the posterior PDF. For a more detailed, introductory exposition of the MH algorithm see [54].

Defining $\Lambda(\boldsymbol{\theta}) = -\log[p(\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta})]$, the posterior PDF equation (1) can be rewritten as

$$p(\boldsymbol{\theta}|\mathbf{z}) = \exp\{-\Lambda(\boldsymbol{\theta})/Z\}, \qquad \text{where} \quad Z = \int \exp\{-\Lambda(\boldsymbol{\theta})\} \, d\boldsymbol{\theta}.$$

In this form, the posterior PDF will be familiar to physicists as the evaluation of Boltzmann factors and partition functions in statistical mechanics. This was the context in which the original version of the Metropolis algorithm was developed [33]. Instead of drawing independent samples from $p(\boldsymbol{\theta}|\mathbf{z})$ directly, loosely speaking (but made more precise below) a random walk in the parameter space is constructed such that the probability for being in a certain region of the space is proportional to the posterior density for that region. In other words, a sample from a Markov chain is generated which has $p(\boldsymbol{\theta}|\mathbf{z})$ as its equilibrium distribution. Starting with an arbitrary initial value of the parameter, say $\boldsymbol{\theta}^{(0)}$, and the evaluation of the posterior PDF at $\boldsymbol{\theta}^{(0)}$, one takes a small random step to another candidate parameter, say $\boldsymbol{\theta}'$, and re-evaluates the posterior PDF. This step is accepted or rejected according to some rule which depends on the values of the posterior PDF at $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{\theta}'$, which guarantees that the so-constructed Markov chain has $p(\boldsymbol{\theta}|\mathbf{z})$ as its equilibrium distribution. The process is then repeated.

More explicitly, starting with an arbitrary initial parameter value $\boldsymbol{\theta}^{(0)}$, a new parameter value, $\boldsymbol{\theta}'$, is generated from a candidate generating, proposal function $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})$, which is symmetric in $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{(0)}$, i.e. $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)}) = q(\boldsymbol{\theta}^{(0)}|\boldsymbol{\theta})$; one example could be a multivariate normal with mean $\boldsymbol{\theta}^{(0)}$ and constant covariance matrix Σ. This candidate point $\boldsymbol{\theta}'$ is accepted with probability

$$\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(0)}) = \min\left\{1, \frac{p(\boldsymbol{\theta}')p(\mathbf{z}|\boldsymbol{\theta}')}{p(\boldsymbol{\theta}^{(0)})p(\mathbf{z}|\boldsymbol{\theta}^{(0)})}\right\}.$$

This means, if the posterior probability at θ' is larger than at $\theta^{(0)}$, the proposed step to θ' is always accepted. But if the step is in a direction with lower posterior probability, then this step is accepted only with a certain probability given by the ratio of the posterior PDFs. If the candidate is accepted, the next state becomes $\theta^{(1)} = \theta'$, otherwise the chain does not move, i.e. $\theta^{(1)} = \theta^{(0)}$.

Hastings [34] generalized this to the following algorithm, where an arbitrary, not necessarily symmetric, candidate generating density can be used. Being in state $\theta^{(n)}$ at time n, a new candidate θ' is accepted with a certain *acceptance probability* $\alpha(\theta'|\theta^{(n)})$, also depending on the current state, $\theta^{(n)}$ given by:

$$\alpha(\theta'|\theta^{(n)}) = \min \left\{ \frac{p(\theta')p(\mathbf{z}|\theta')q(\theta^{(n)}|\theta')}{p(\theta^{(n)})p(\mathbf{z}|\theta^{(n)})q(\theta'|\theta^{(n)})}, 1 \right\}$$

if $(p(\theta^{(n)})p(\mathbf{z}|\theta^{(n)})q(\theta'|\theta^{(n)})) > 0$, and $\alpha(\theta'|\theta^{(n)}) = 1$ otherwise. For good efficiency a multivariate normal distribution is used for $q(\theta|\theta^{(n)})$. The steps of the MH algorithm are therefore:

Step 0:   Start with an arbitrary value $\theta^{(0)}$.
Step $n + 1$:   Generate θ' from $q(\theta|\theta^{(n)})$ and u from $U(0, 1)$;
            If $u \leqslant \alpha(\theta'|\theta^{(n)})$ set $\theta^{(n+1)} = \theta'$ (acceptance);
            If $u > \alpha(\theta'|\theta^{(n)})$ set $\theta^{(n+1)} = \theta^{(n)}$ (rejection).

The MH algorithm thus generates a Markov chain, i.e. a sequence of random variables $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \ldots$ such that at time n, $\theta^{(n+1)}$ is sampled from a distribution $p(\theta^{(n+1)}|\theta^{(n)})$ that depends only on the current state, but not on the further history of the chain, $\theta^{(0)}, \ldots, \theta^{(n-1)}$. $p(.|.)$ is called the transition kernel of the Markov chain. It is assumed that the Markov chain is time-homogeneous, i.e. the transition kernel does not depend on n.

The usual concern of Markov chain theory is to determine conditions under which there exists an invariant distribution, say $\pi(\theta)$, and conditions under which iterations of the transition kernel converge to the invariant (or stationary or equilibrium) distribution. The invariant distribution $\pi(\theta)$ satisfies

$$\pi(\mathrm{d}\theta) = \int p(\theta|\theta^{(n)})\pi(\theta)\,\mathrm{d}\theta$$

which means that once the chain reaches a stage where π is the distribution of the chain, the chain retains this distribution for all subsequent stages. MCMC methods turn this theory around: the invariant distribution $\pi(\theta)$ is known (the posterior PDF $p(\theta|\mathbf{z})$ up to a normalization constant), but the transition kernel is unknown. A sufficient condition [55] for a distribution π to be the invariant distribution is the 'detailed balance equation', or the 'time reversibility' condition,

$$\pi(\theta^{(n)})p(\theta^{(n+1)}|\theta^{(n)}) = \pi(\theta^{(n+1)})p(\theta^{(n)}|\theta^{(n+1)})$$

for all possible states $\theta^{(n)}$, $\theta^{(n+1)}$, which can be interpreted as saying that the rate at which the system moves from $\theta^{(n)}$ to $\theta^{(n+1)}$ when in equilibrium, $\pi(\theta^{(n)})p(\theta^{(n+1)}|\theta^{(n)})$, is the same as the rate at which it moves from $\theta^{(n+1)}$ to $\theta^{(n)}$, for $\pi(\theta^{(n+1)})p(\theta^{(n)}|\theta^{(n+1)})$.

Thus, if one wants to construct a transition kernel such that the corresponding Markov chain has a given $\pi(\theta)$ as its invariant distribution, the idea is to start with a candidate generating density $q(\theta|\theta^{(n)})$ that may depend on the current state $\theta^{(n)}$. If q itself satisfies detailed balance for all $\theta^{(n)}$, $\theta^{(n+1)}$, the search for a transition kernel would be over. But if, for instance, for some $\theta^{(n+1)}$, $\theta^{(n)}$,

$$\pi(\theta^{(n)})q(\theta^{(n+1)}|\theta^{(n)}) > \pi(\theta^{(n+1)})q(\theta^{(n)}|\theta^{(n+1)}) \tag{5}$$

the process moves from $\boldsymbol{\theta}^{(n)}$ to $\boldsymbol{\theta}^{(n+1)}$ too often, and from $\boldsymbol{\theta}^{(n+1)}$ to $\boldsymbol{\theta}^{(n)}$ too rarely. To correct this, one reduces the number of moves from $\boldsymbol{\theta}^{(n)}$ to $\boldsymbol{\theta}^{(n+1)}$ by introducing an acceptance probability $\alpha(\boldsymbol{\theta}^{(n+1)}|\boldsymbol{\theta}^{(n)})$. Thus, transitions from $\boldsymbol{\theta}^{(n)}$ to $\boldsymbol{\theta}^{(n+1)}$ are made according to

$$p(\boldsymbol{\theta}^{(n+1)}|\boldsymbol{\theta}^{(n)}) = q(\boldsymbol{\theta}^{(n+1)}|\boldsymbol{\theta}^{(n)})\alpha(\boldsymbol{\theta}^{(n+1)}|\boldsymbol{\theta}^{(n)}),$$

where $\alpha(\boldsymbol{\theta}^{(n+1)}|\boldsymbol{\theta}^{(n)})$ still needs to be determined from the detailed balance equations. As moves from $\boldsymbol{\theta}^{(n+1)}$ to $\boldsymbol{\theta}^{(n)}$ are not made often enough, the acceptance probability $\alpha(\boldsymbol{\theta}^{(n)}|\boldsymbol{\theta}^{(n+1)})$ should be set equal to 1. The detailed balance equations then give

$$\pi(\boldsymbol{\theta}^{(n)})q(\boldsymbol{\theta}^{(n+1)}|\boldsymbol{\theta}^{(n)})\alpha(\boldsymbol{\theta}^{(n+1)}|\boldsymbol{\theta}^{(n)}) = \pi(\boldsymbol{\theta}^{(n+1)})q(\boldsymbol{\theta}^{(n)}|\boldsymbol{\theta}^{(n+1)})\alpha(\boldsymbol{\theta}^{(n)}|\boldsymbol{\theta}^{(n+1)})$$
$$= \pi(\boldsymbol{\theta}^{(n+1)})q(\boldsymbol{\theta}^{(n)}|\boldsymbol{\theta}^{(n+1)})$$

and thus

$$\alpha(\boldsymbol{\theta}^{(n+1)}|\boldsymbol{\theta}^{(n)}) = \frac{\pi(\boldsymbol{\theta}^{(n+1)})q(\boldsymbol{\theta}^{(n)}|\boldsymbol{\theta}^{(n+1)})}{\pi(\boldsymbol{\theta}^{(n)})q(\boldsymbol{\theta}^{(n+1)}|\boldsymbol{\theta}^{(n)})},$$

the acceptance probability of the MH algorithm. The acceptance probability is thus introduced to ensure that both sides in equation (5) are in balance, i.e. that the MH transition kernel satisfies reversibility. As shown in [55], the MH kernel has $\pi(\boldsymbol{\theta})$ as its invariant density. Convergence towards the invariant distribution occurs under mild regularity conditions [55].

Note that the MH algorithm does not require the normalization constant of the target density. The outcomes from the MH algorithm can be regarded as a sample from the invariant density only after a certain *burn-in* period. Thus, after running the Markov chain for a certain 'burn-in' period, these (correlated) samples can be regarded as samples from the limiting distribution, provided that the Markov chain has reached convergence. Despite their correlations, the ergodic theorem guarantees that the sample average, equation (4), is still a consistent estimate of the posterior mean equation (2). Various methods to assess convergence, i.e. methods used for establishing whether an MCMC algorithm has converged and whether its output can be regarded as samples from the target distribution of the Markov chain, have been developed and implemented [56].

Computational complexity and rates of convergence of MCMC methods are still active areas of research [57, 58]. Roberts *et al* [59] show geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. In high-dimensional problems of Bayesian image reconstruction, where each pixel plays the role of a parameter, Frigessi *et al* [60] verify convergence of the Gibbs sampler to the stationary distribution in polynomial time under certain conditions.

The efficiency of the MH algorithm depends heavily on the choice of the proposal density. The random walk MH algorithm, for instance, suffers from slow convergence and poor mixing when applied to complex statistical models or multi-modal problems. The MH algorithm can often be improved by a careful tuning of the proposal density. The closer the proposal is to the target distribution, the faster convergence will be accomplished. This link between the closeness of the proposal to the stationary distribution, and speed of convergence has also been substantiated by Holden [61]. The specification of a suitable proposal density may require a careful analysis of the target density, such as first-order derivation to find the mode, and an evaluation of the second-order derivatives at the mode to construct the covariance matrix of a Gaussian proposal. In this paper, we pursue an alternative strategy that avoids the need for preliminary analytic investigations to design custom-made proposals. We dynamically alter the proposal distribution based on information from the chain's history. The approach, called *pilot adaptation*, is to perform a separate pilot run to gain insight into the target density, and then tune the proposal accordingly for the successive runs. Such adaptation can be iterated, but allowing it infinitely often will destroy the Markovian property of the chain and thereby often

compromise the stationarity of the chain and the consistency of sample path averages ([62], see [63] for an example). In our study we conducted a single *pilot run* using independent normal distributions for the proposal density of each parameter. As explained below, the information from the pilot run was then used to generate a multivariate normal distribution function, from which all parameters were then sampled. The general choice of a multivariate normal proposal density is based on the Laplace approximation [64] which states that the posterior distribution is approximately Gaussian with mean equal to the posterior mode and covariance matrix equal to the inverse of the Hessian matrix of the logarithm of the posterior density evaluated at the posterior mode. Thus, a multivariate normal distribution will in general be close to the posterior density and, furthermore, fast and easy to sample from. This makes the multivariate normal an efficient choice for the candidate generating function.

4. Examples of coalescing binary signals detected by a laser interferometer

The detector output, $z(t)$, is the sum of the gravitational radiation signal, $s(t, \boldsymbol{\theta})$, that depends on unknown parameters $\boldsymbol{\theta}$, and the noise $n(t)$, namely $z(t) = s(t, \boldsymbol{\theta}) + n(t)$ for $t \in [0, t_u]$. We assume that the noise is Gaussian with mean zero and known one-sided power spectral density $S_h(f)$. The signal-to-noise ratio (SNR) of the *detected* signal is $\text{SNR} = \sqrt{4 \int_0^\infty \frac{\tilde{s}^*(f)\tilde{s}(f)}{S_h(f)} \, \mathrm{d}f}$, with $\tilde{s}(f) = \int_{-\infty}^\infty s(t) \, \mathrm{e}^{2\pi \mathrm{i} f t} \, \mathrm{d}t$ the Fourier transform of the function $s(t)$ [13]. The *likelihood* is given by

$$p(\mathbf{z}|\boldsymbol{\theta}) = K \exp[2\langle z, s(\boldsymbol{\theta}) \rangle - \langle s(\boldsymbol{\theta}), s(\boldsymbol{\theta}) \rangle], \tag{6}$$

where K is a constant and $\langle a, b \rangle = \int_{-\infty}^\infty \mathrm{d}f \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_h(|f|)}$ the inner product of two functions a, b [49].

The signals were generated using the code *inject*, created by T Creighton of Caltech and implemented in LAL [32]. The noise level was kept constant for all of the tests used in our study, while the signal-to-noise ratio was varied by changing the distance to our fictitious source. Also, for this study the spectral density of the noise was white extending from 40 Hz through the sampling frequency of 2048 Hz. The datasets were 64 s in length. For this study both the signals and the detection templates were calculated to 2.0 post-Newtonian order; the exact form of the signal is presented elsewhere [65, 66].

Our initial research goal, presented in this paper, is to demonstrate the usefulness of the MH-MCMC technique for estimating parameters from coalescing binary signals detected by laser interferometric antennas. The MCMC routine is designed to eventually operate within LAL [32]. The core of the inspiral event programme comes from the LAL programme 'findchirp' [32], developed by Duncan Brown and Patrick Brady (University of Wisconsin, Milwaukee). We have written our MCMC routine so that it will call the same inspiral templates that findchirp does. As new and higher order templates are developed, they will automatically be incorporated and used through findchirp. The MCMC routine generates chains for the following parameters: the binary star masses m_1, m_2, and the effective distance to the source D. The coalescence time t_c and phase ϕ_c are found through the FFT in findchirp, and are not part of the MCMC output in this present study. The MH routine calculates the likelihood, equation (6), from the data and the template for the parameter values being used at that point in the Markov chain.

With the MH code it was found that the form of the gravitational radiation signal makes the search process difficult. There are competing issues that need to be satisfied. When the code is doing its initial burn-in search for the signal it is necessary to adequately sample the entire parameter space, and the programme needs to move with relatively large steps. However, after

the chain has converged the programme needs to take relatively small steps; these small step sizes produce a larger acceptance probability for the Markov chain, and thereby make for a more efficient sampling in order to efficiently generate the probability distribution functions. The most efficient MCMC code (post burn-in) samples candidate parameter values from a distribution that closely resembles the posterior PDF of the parameters. Through much work, a strategy that satisfies both criteria was derived.

In our study a single, initial pilot run was conducted, with normal distributions for the candidate generating functions. At step n in the Markov chain, with ith parameter value $\theta_i^{(n)}$, the next candidate parameter $\theta_i^{(n+1)}$ is generated by sampling from a normal distribution

$$\frac{1}{\sqrt{2\pi}\,\epsilon_i}\, \mathrm{e}^{-\frac{(\theta_i-\theta_i^{(n)})^2}{2\epsilon_i^2}}, \tag{7}$$

where the factor ϵ_i defines the spread of the distribution for each parameter. We found the most efficient strategy was for the ϵ_i values to change with iteration number in the Markov chain. In the initial part of the run one needs a relatively large value, $\epsilon_{i\text{-big}}$, for sufficient sampling of the parameter space. After *burn-in* and convergence of the chain towards the actual parameter values, it is necessary to have a small value, $\epsilon_{i\text{-small}}$, for efficient sampling to generate the kernel densities. In our MCMC run we used

$$\epsilon_i = \epsilon_{i\text{-small}} + \epsilon_{i\text{-big}} \times \exp(-n/N), \tag{8}$$

where n is the location in the Markov chain, while N is a *time constant* that is chosen in order to optimize efficiency (typically 5×10^3). At the start of the chain a relatively large value of ϵ_i is used, namely $\epsilon_{i\text{-big}} + \epsilon_{i\text{-small}}$. As the chain progresses the value of ϵ_i decreases exponentially towards $\epsilon_{i\text{-small}}$. This process allows the chain to sample a large region initially, and after burn-in it can effectively sample the smaller region about the parameter estimate. The most useful values we used for the mass and distance parameters were: $\epsilon_{m1\text{-big}} = \epsilon_{m2\text{-big}} = 0.5 M_\odot$, $\epsilon_{m1\text{-small}} = \epsilon_{m2\text{-small}} = 0.0015 M_\odot$, $\epsilon_{D\text{-big}} = 10.0$ Mpc, and $\epsilon_{D\text{-small}} = 1.0$ kpc.

The *a priori* distributions were all uniform in this study, in order to keep this investigation simple; the masses for the compact objects were restricted to the range of $1.1\ M_\odot$ to $10.0\ M_\odot$. The amplitude of the gravitational radiation falls off as $1/D$, while the number of potential sources grows as D^3. An accurate prior for the distance parameter would take this into account. When we ultimately apply this technique to real data from gravitational radiation interferometers the space and mass distributions of the neutron stars and black holes will be based on astronomical information [67, 68]. For simplicity in this demonstration study we take a uniform prior for D on [0, 1000 Mpc].

From the pilot run samples for each parameter $\boldsymbol{\theta} = (m_1, m_2, D)$ a cross-correlation matrix, $\mathbf{C} = (C_{ij})$, was calculated. The cross-correlation is a standard method of estimating the degree to which two series are correlated. Consider the two MCMC output series $\theta_i^{(n)}$ and $\theta_j^{(n)}$, where $n = 1, 2, 3, \ldots, N$. The cross-correlation matrix we used was defined as

$$C_{ij} = \frac{\sum_{i=1}^{N}\left(\theta_i^{(n)} - \hat{\theta}_i\right)\left(\theta_j^{(n)} - \hat{\theta}_j\right)}{\sqrt{\sum_{n=1}^{N}\left(\theta_i^{(n)} - \hat{\theta}_i\right)^2}\sqrt{\sum_{n=1}^{N}\left(\theta_j^{(n)} - \hat{\theta}_j\right)^2}}, \tag{9}$$

where $\hat{\theta}_i$ and $\hat{\theta}_j$ are the means of the sample averages, as given by equation (4).

After this single pilot run an optimized Metropolis–Hastings code was created using a multivariate-normal distribution for the candidate generating function $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)})$, where the mean parameter values were just the last values accepted into the chain, $\boldsymbol{\theta}^{(n)}$. Specifically,

$$q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\, \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(n)})'\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(n)})}, \tag{10}$$

where d is the number of parameters. The covariance matrix of the multivariate normal
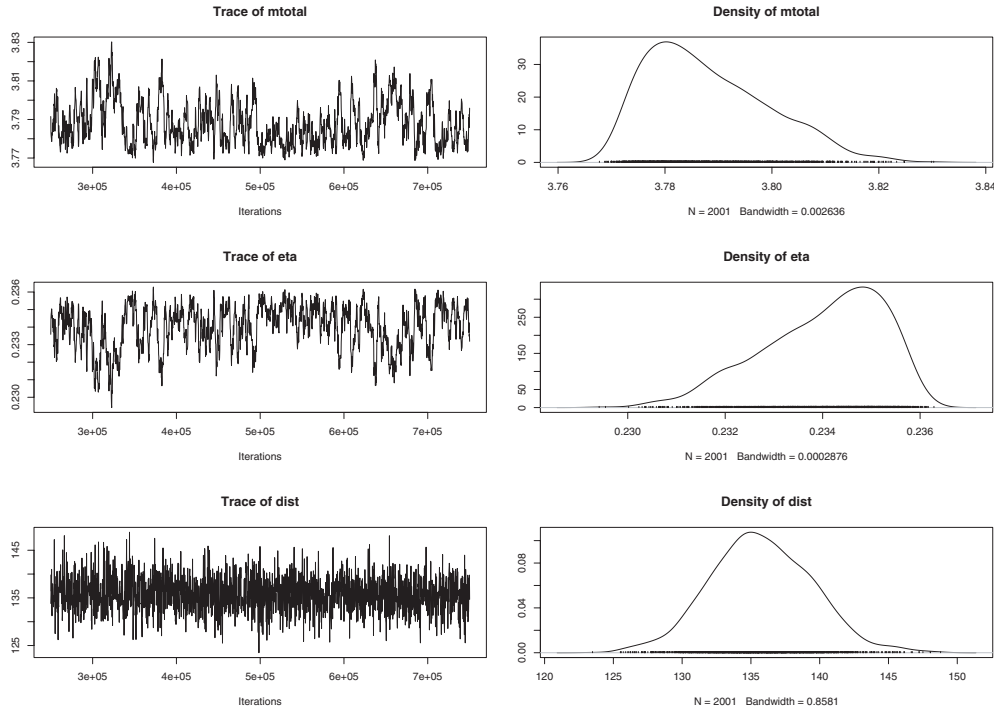
Figure 1. The trace plots and kernel densities for the parameters m_t (in units of solar masses, $M_\odot$), η and D, generated from the MH routine with SNR = 14.

proposal distribution is a scaled version of **C**: for diagonal elements we had $\Sigma_{ii} = \epsilon_i\epsilon_i C_{ii} = C_{ii}(\epsilon_{i\text{-small}} + \epsilon_{i\text{-big}} \times \exp(-n/N))^2$, while for the off-diagonal elements we used $\Sigma_{ij} = C_{ij}\epsilon_{i\text{-small}}\epsilon_{j\text{-small}} \times (1 - \exp(-n/N))$. We used the same numerical values for the ϵ' as in the pilot run, and again n is the location in the Markov chain, while N is the *time constant* that is chosen in order to optimize efficiency (again 5×10^3). The prior ranges for the parameters were the same as well. For the multivariate-normal run we again wanted an efficient burn-in, where the parameter space is effectively sampled, with large ϵ_i and zero values for the off-diagonal elements of Σ_{ij}. As the chain progresses the value of ϵ_i decreases exponentially towards $\epsilon_{i\text{-small}}$, and the covariance matrix approaches $\Sigma_{ij} = C_{ij}\epsilon_{i\text{-small}}\epsilon_{j\text{-small}}$. After burn-in the multivariate-normal generating function can effectively sample the smaller region about the parameter estimate.

The number of iterations for each chain in this study was 7.5×10^5. The lower frequency for the generated signal was 40 Hz. The response function for the interferometer in this study was assumed to be flat, and the noise spectrum was white. For 64 s datasets, sampled at 2048 Hz, the 7.5×10^5 MCMC iterations took $\sim$15 h to complete on a 600 MHz pc. The first 2.5×10^5 points were discarded for the burn-in, and the subsequent data were sampled at every 250th point. The PDFs were generated from the remaining samples.

For our first example we studied a binary system with masses $m_1 = 1.4M_\odot$, $m_2 = 2.4M_\odot$. The templates for the gravity wave signals depend on the total mass $m_t = m_1 + m_2$, and the mass ratio $\eta = m_1 m_2/m_t^2$, and we express the MCMC output in terms of these mass parameters. In the MCMC code we sample m_1 and m_2, and then calculate m_t and η from these values. For this example we have $m_t = 3.8M_\odot$ and $\eta = 0.233$. We set the lower frequency used for the templates at 75 Hz. An example of the MH routine output is displayed in figure 1.
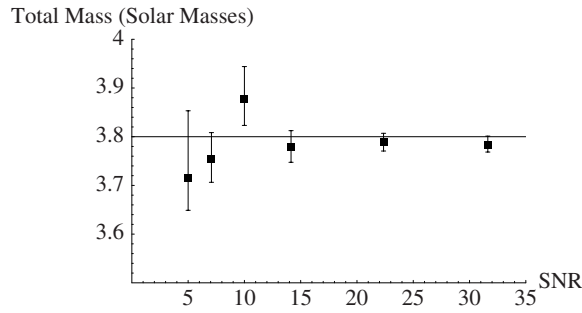
Total Mass (Solar Masses)

Figure 2. Estimate of the total mass, m_t, versus SNR. The actual total mass for the signal is $m_t = 3.8 M_\odot$. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of m_t (which gives a 95% posterior credibility interval for m_t).
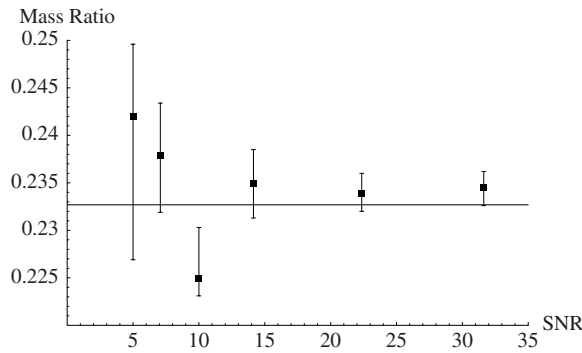
Mass Ratio

Figure 3. Estimate of the mass ratio, η, versus SNR. The actual mass ratio for the signal is $\eta = 0.233$. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of η (which gives a 95% posterior credibility interval for η).

The trace plots and kernel densities for the parameters m_t (in units of solar masses, $M_\odot$), η, and D, generated from the MH routine with SNR $= 14$ are shown. The functional character of the signal makes m_t and η strongly negatively correlated, which is also observable in these chains. For the simulations we kept the noise level constant, but varied the SNR by adjusting the distance to the source. In figure 2 we plot the estimate of the total mass, m_t, versus SNR. In figure 3 we plot the estimate of the mass ratio, η, versus SNR. As expected, the widths of the distributions (and hence the uncertainty in the estimate of the parameter value) increase as the SNR decreases.

In another example we looked at a signal from a more massive system, with masses $m_1 = 3.0 M_\odot$, $m_2 = 4.5 M_\odot$. For this example we have $m_t = 7.5 M_\odot$ and $\eta = 0.24$. We set the lower frequency used for the templates at 60 Hz. An example of the MH routine output is displayed in figure 4. The trace plots and kernel densities for the parameters m_t, η and D, generated from the MH routine with SNR $= 7$, are shown. In figure 5 we plot the estimate of the total mass, m_t, versus SNR, while in figure 6 we again plot the estimate of the mass ratio, η, versus SNR.

For signals with SNR < 5 the MCMC programme would not converge in a reasonable amount of time, or 2.5×10^5 iterations. This is consistent with the signal being too small, and the MCMC routine being unable to find it. We also ran the programme when there was
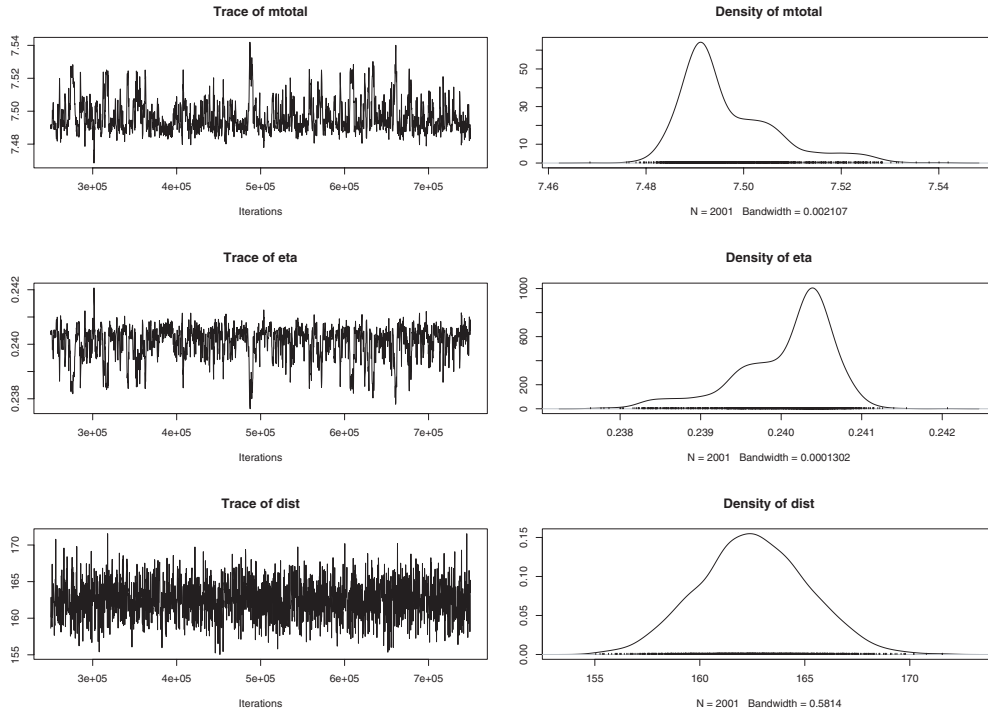
Figure 4. The trace plots and kernel densities for the parameters m_t (in units of solar masses, $M_\odot$), η and D, generated from the MH routine with SNR = 7.
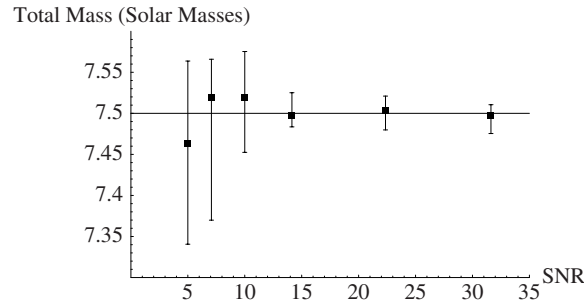

Figure 5. Estimate of the total mass, m_t, versus SNR. The actual total mass for the signal is $m_t = 3.8M_\odot$. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of m_t (which gives a 95% posterior credibility interval for m_t).

no signal present, and again the output chains oscillated around their allowed prior regions, indicative of no convergence and no signal detected. Hence there was no false positive.

5. Discussion

In the study presented in this paper we have shown that a MH routine offers hope for estimating the parameter values from a gravitational radiation signal created by a coalescing binary system, and also for generating the associated PDFs for the parameters. We have
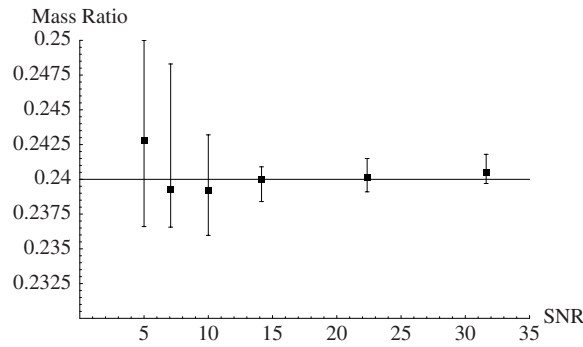
Figure 6. Estimate of the mass ratio, η, versus SNR. The actual mass ratio for the signal is $\eta = 0.233$. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of η (which gives a 95% posterior credibility interval for η).

demonstrated that a MH routine can find (synthesized) gravitational radiation events, determine the parameters of the coalescing binary, and generate summary statistics for the parameters. We have confined this study to relatively simple (2.0 post-Newtonian) signals, but in principle this MH routine can easily use more complex templates.

The success of any gravitational radiation search for binary inspiral events is critically dependent on the accuracy of the theoretically generated templates. In our study we use these theoretical waveforms to improve the efficiency of our MCMC routine. An efficient MCMC programme will adequately sample the parameter space, not get stuck in local maximum of the posterior PDF, and find the global extrema in a timely fashion. The ability to accomplish this task depends on both the functional character of the candidate generating function, and how long (in number of programme iterations) one is willing to let the burn-in of the MCMC go. Through code optimization, and a burn-in limit of 2.5×10^5, we were always able to *find* the signal when SNR > 5. However, when the candidate generating function was not tuned, the chain often converged to incorrect parameter values (a local extremum). Hence the probability of a false detection depends on the burn-in time and the candidate generating PDF. By carefully tuning the multivariate-normal candidate generating function we were able to achieve good success. We ultimately desire to apply this technique to real data from interferometric detectors. By using synthesized data from theoretically generated signals we can optimize our routine, but the success of the code ultimately depends on the accuracy of the theoretical signals.

Our current research effort is to extend the work presented here to more complex signal waveforms. As the number of parameters increases the search for signals by a systematic march over a template grid will become prohibitively difficult. The MH routine, on the other hand, may be able to allow signal detection and parameter estimation for complex signals. We are currently examining signals from black hole—black hole ($1M_\odot$ to $20M_\odot$ for each) inspirals; the total waveform encompasses the final plunge and subsequent ringing of the black hole [41, 42]. We are also expanding our study to account for the spin of each compact object [37–40]. Finally, we hope to ultimately apply MCMC techniques to the coherent response of a network of detectors located around the world; in this situation the additional parameters are the location of the source in the sky and the polarization of the gravitational radiation.

Present gravitational radiation searches for binary inspiral coalescence make use of a grid in parameter space, whereby the templates associated with a given grid point are compared via matched filtering with the data. At the most basic level, the MH-MCMC routine we describe

in this paper could supplement this type of search. For any time stretches of data where *events* were found via the grid search, the MCMC could then subsequently examine them. The MCMC would provide PDFs and statistics for the parameters associated with an event; this would not be an easy task to do from the grid-based search alone. In the future, when the search templates become far more complicated due to spin interactions, final mass plunge and black-hole ringing, it is not clear how one would even construct an efficient grid on the parameter space. Our long-term goal would be to create a fast and efficient code, running on multiple processors, that would send all of the data through the MCMC routine.

MCMC techniques have been demonstrated as a method that can handle large numbers of parameters [24]. MCMC methods may prove promising for dealing with complicated binary inspiral signals that are observed with a network of detectors. In the study presented in this paper we have demonstrated that a MH routine offers great potential for binary inspiral gravitational radiation signal analysis.

Acknowledgments

References

[1] Abramovici A, Althouse W E, Drever R W P, Gürsel Y, Kawamura S, Raab F J, Shoemaker D, Sievers L, Spero R E, Thorne K S, Vogt R E, Weiss R, Whitcomb S E and Zucker M E 1992 *Science* **256** 325
[2] Barish B C 1997 *Gravitational Wave Detection* ed K Tsubono, M-K Fujimoto and K Kurodo (Tokyo: Universal Academic) pp 155–61
[3] Tsubono K 1997 *Gravitational Wave Detection* ed K Tsubono, M-K Fujimoto and K Kurodo (Tokyo: Universal Academic) pp 183–91
[4] Hough J 1997 *Gravitational Wave Detection* ed K Tsubono, M-K Fujimoto and K Kurodo (Tokyo: Universal Academic) pp 175–82
[5] Caron B *et al* 1996 *Nucl. Phys. Suppl.* **48** 107
[6] Brillet A 1997 *Gravitational Wave Detection* ed K Tsubono, M-K Fujimoto and K Kurodo (Tokyo: Universal Academic) pp 163–73
[7] Einstein A 1916 *Preuss. Akad. Wiss. Berlin, Sitzungsberichte der physikalisch-mathematischen Klasse* p 688
[8] Taylor J and Weisberg J 1889 Further experimental tests of relativistic gravity using the binary pulsar PSR 1913 + 16 *Astrophys. J.* **345** 434
[9] Thorne K S 1987 *300 Years of Gravitation* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) p 330
[10] Markovic' D 1993 *Phys. Rev.* D **48** 4738
[11] Schutz B F 1996 *Nature* **323** 310
[12] Cutler C and Flanagan E E 1994 *Phys. Rev.* D **49** 2658
[13] Finn L S 1996 *Phys. Rev.* D **53** 2878
[14] Cutler C *et al* 1993 *Phys. Rev. Lett.* **70** 2984
[15] Hughes S 2002 *Phys. Rev.* D **66** 102001
[16] Vallisneri M 2002 *Preprint* gr-qc/0202037
[17] Flanagan E E and Hughes S A 1998 *Phys. Rev.* D **57** 4535
[18] Flanagan E E and Hughes S A 1998 *Phys. Rev.* D **57** 4566
[19] Gregory P C and Loredo T J 1992 *Astrophys. J.* **398** 146
[20] Loredo T J 1992 The promise of Bayesian inference for astrophysics *Statistical Challenges in Modern Astronomy* ed E D Feigelson and G J Babu (New York: Springer) pp 275–97
[21] Loredo T J 1996 The return of the prodigal: Bayesian inference for astrophysics *Bayesian Statistics* vol 5 ed J Bernardo *et al* (Oxford: Oxford University Press)

[22] Loredo T J 1999 Computational technology for Bayesian inference *ADASS VIII* ed R Crutcher and D Mehringer (San-Francisco: Astronomical Society of the Pacific)
[23] Finn L S 1997 *Preprint* gr-qc/9709077
[24] Gilks W R, Richardson S and Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall)
[25] Gamerman D 1997 *Markov Chain Monte Carlo* (London: Chapman and Hall)
[26] Chen M H, Shao Q M and Ibrahim J G 2000 *Monte Carlo Methods in Bayesian Computation* (New York: Springer)
[27] Christensen N L and Meyer R 1998 *Phys. Rev.* D **58** 082001
[28] Gilks W R, Best N G and Chan K 1995 *Appl. Stat.* **44** 445
[29] Christensen N and Meyer R 2001 *Phys. Rev.* D **64** 022001
[30] Spiegelhalter D J, Thomas A, Best N and Gilks W R 1996 *BUGS 0.5 Bayesian Inference Using Gibbs Sampling Manual (Version ii)* (Cambridge: MRC Biostatistics Unit)
[31] LIGO Scientific Collaboration 2002 www.ligo.org
[32] LAL, LIGO/LSC Algorithm Library Home Page 2002 http://www.lsc-group.phys.uwm.edu/lal/index.html
[33] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[34] Hastings W K 1970 *Biometrika* **57** 97
[35] Christensen N, Meyer R, Knox L and Luey B 2001 *Class. Quantum Grav.* **18** 2677
[36] Knox L, Christensen N and Skordis C 2001 *Astrophys. J. Lett.* **563** L95
[37] Apostolatos T A 1995 *Phys. Rev.* D **52** 605
[38] Apostolatos T A 1996 *Phys. Rev.* D **54** 2421
[39] Apostolatos T A 1996 *Phys. Rev.* D **54** 2438
[40] Chronopoulos A E and Apostolatos T A 2001 *Phys. Rev.* D **64** 042003
[41] Damour T, Iyer B R and Sathyaprakash B S 2001 *Phys. Rev.* D **63** 044023
[42] Damour T, Iyer B R and Sathyaprakash B S 2002 *Phys. Rev.* D **66** 027502
[43] Buonanno A and Damour T 1999 *Phys. Rev.* D **59** 084006
[44] Buonanno A and Damour T 2000 *Phys. Rev.* D **62** 064015
[45] Bose S 2002 *Class. Quantum Grav.* **19** 1437–42
[46] Pai A, Bose S and Dhurandhar S 2002 *Class. Quantum Grav.* **19** 1477
[47] Finn L S 1992 *Phys. Rev.* D **46** 5236
[48] Finn L S 1997 *Preprint* gr-qc/9709077
[49] Finn L S and Chernoff D 1993 *Phys. Rev.* D **47** 2198
[50] Mohanty S D 1998 *Phys. Rev.* D **57** 630
[51] Nicholson F and Vecchio A 1998 *Phys. Rev.* D **57** 4588
[52] Balasubramanian R and Dhurandhar S V 1998 *Phys. Rev.* D **57** 3408
[53] Dongarra J and Sullivan F 2000 *Comput. Sci. Eng.* **2** 22
[54] Chib S and Greenberg E 1995 *Am. Stat.* **49** 327
[55] Tierney L 1994 *Ann. Stat.* **22** 1701–62
[56] Best N G, Cowles M K and Vines S K 1995 *CODA Manual Version 0.30* (Cambridge: MRC Biostatistics Unit)
[57] Roberts G O and Sahu S K 2001 *J. Comput. Graph. Stat.* **10** 216
[58] Roberts G O and Rosenthal J S 1999 *J. R. Stat. Soc.* B **61** 643
[59] Roberts G O and Tweedie R L 1996 *Biometrika* **83** 95
[60] Frigessi A, Martinelli F and Stander J 1997 *Biometrika* **84** 1
[61] Holden L 2000 Adaptive chains *Preprint* http://www.statslab.cam.ac.uk/mcmc
[62] Gilks W R, Roberts G O and Sahu S K 1998 *J. Am. Stat. Assoc.* **93** 1045
[63] Gelfand A E and Sahu S K 1994 *J. Comput. Graph. Stat.* **3** 261–76
[64] Laplace P S 1986 *Stat. Sci.* **1** 364
[65] Blanchet L, Iyer B R, Will C M and Wiseman A G 1996 *Class. Quantum Grav.* **13** 575
[66] Tanaka T and Tagoshi H 2000 *Phys. Rev.* D **62** 082001
[67] Belczynski K, Kalogera V and Bulik T 2002 *Astrophys. J.* **572** 407
[68] Kim C, Kalogera V and Lorimer D R 2002 *Preprint* astro-ph/0207408