INSTITUTE OF PHYSICS PUBLISHING CLASSICAL AND QUANTUM GRAVITY

Class. Quantum Grav. **21** (2004) S1655–S1665 PII: S0264-9381(04)78801-2

Estimating the parameters of gravitational waves from neutron stars using an adaptive MCMC method

Richard Umstätter[1], Renate Meyer[1], Réjean J Dupuis[2], John Veitch[2], Graham Woan[2] and Nelson Christensen[3]

[1] Department of Statistics, University of Auckland, Auckland, New Zealand
[2] Department of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, UK
[3] Physics and Astronomy, Carleton College, Northfield, MN 55057, USA

E-mail: richard@stat.auckland.ac.nz, meyer@stat.auckland.ac.nz, rejean@astro.gla.ac.uk, jveitch@astro.gla.ac.uk, graham@astro.gla.ac.uk and nchriste@carleton.edu

Received 3 April 2004
Published 24 September 2004
Online at stacks.iop.org/CQG/21/S1655
doi:10.1088/0264-9381/21/20/008

Abstract

We present a Bayesian Markov chain Monte Carlo technique for estimating the astrophysical parameters of gravitational radiation signals from a neutron star in laser interferometer data. This computational algorithm can estimate up to six unknown parameters of the target, including the rotation frequency and frequency derivative, using reparametrization, delayed rejection and simulated annealing. We highlight how a simple extension of the method, distributed over multiple computer processors, will allow for a search over a narrow frequency band. The ultimate goal of this research is to search for sources at known locations, but uncertain spin parameters; an example would be SN1987A.

PACS numbers: 04.80.Nn, 02.70.Uu, 06.20.Dk

1. Introduction

Rapidly rotating neutron stars could be an important source of gravitational wave signals. Several mechanisms have been proposed that would cause them to emit quasi-periodic gravitational waves [1, 2].

Interferometric gravitational wave detectors are now operating in numerous locations around the world [3–6], and much work has gone into the development of dedicated search algorithms for these signals. Radio observations can provide the sky location, rotation frequency and spin-down rate of known pulsars, and this knowledge simplifies the analysis. This was the case for the recent search for a signal from PSR J1939+2134 [7]. When the

position and phase evolution of a source are not known, all-sky hierarchical strategies are required, and these have huge computational requirements [8, 9].

Here we concentrate on the search for a gravitational wave signal from a known location, but where spin parameters of the rotating neutron star are not well known (but within a narrow band). SN1987A is a good example of a poorly parametrized source for which the sky location is known, but where there are large uncertainties in the frequency and spin-down parameters of a putative neutron star [10]. In particular, we consider a search with six unknown parameters: the gravitational wave amplitude h_0, the polarization angle ψ (which depends on the position angle of the spin axis in the plane of the sky), the phase of the signal at a fiducial time ϕ_0, the inclination of the spin axis with respect to the line-of-sight ι and the deviations (from reference values) of the signal frequency Δf, and of the frequency derivative $\Delta \dot{f}$.

We use a Bayesian Markov chain Monte Carlo (MCMC) technique for this analysis as MCMC methods have been applied successfully to similar problems involving large numbers of parameters [11]. In a previous study [14], we used a Metropolis–Hastings (MH) algorithm [12, 13] for a similar search, but with only five parameters ($\Delta \dot{f}$ being absent). When the frequency derivative $\Delta \dot{f}$ is included in the basic MH routine of [14], the large correlation between Δf and $\Delta \dot{f}$ makes the parameter search difficult, and the basic MH algorithm becomes inefficient. In order to adequately sample the parameter space, we implemented a combination of three different strategies for accelerating convergence of Markov chains: reparametrization, the delayed rejection method of Tierney and Mira [15] (which is an adaptive version of the MH algorithm) and simulated annealing [16] (which is a Monte Carlo approach to global optimization). The parameter Δf is highly correlated with $\Delta \dot{f}$, and a strong correlation also exists between h_0 and $\cos\iota$. An initial transformation of these variables to near orthogonality yields a more tractable parameter space that is more effectively sampled.

The heterodyne manipulation of the data used in this study is identical to that presented (by two of us) in an end-to-end robust Bayesian method of searching for periodic signals in gravitational wave interferometer data [17], and is also described in [7]. A brief summary of this heterodyne technique is given in section 2. Our delayed rejection method, as well as the reparametrization strategy, is presented in section 3. In section 4 we present the results of this study, using synthesized signals, for this six-parameter problem. A brief discussion of the long-term goals for this work is presented in section 5.

2. The gravitational wave signal

Gravitational waves from spinning neutron stars are expected to be weak at the Earth, so long integration periods are necessary to extract the signal. It is therefore important to take proper account of the antenna patterns of the detectors and the Doppler shift due to the motion of the Earth.

As in previous studies [7, 14, 17] we consider the signal expected from a non-precessing triaxial neutron star. The gravitational wave signal from such an object is at twice its rotation frequency, $f_\mathrm{s} = 2f_\mathrm{r}$, and we characterize the amplitudes of each polarization with overall strain factor, h_0. The measured gravitational wave signal will also depend on the antenna patterns of the detector for the 'cross' and 'plus' polarizations, $F_{\times,+}$, giving a signal

$$s(t) = \tfrac{1}{2}F_+(t;\psi)h_0(1+\cos^2\iota)\cos\Psi(t) + F_\times(t;\psi)h_0\cos\iota\sin\Psi(t). \tag{1}$$

A simple slowdown model provides the phase evolution of the signal as

$$\Psi(t) = \phi_0 + 2\pi\left[f_\mathrm{s}(T-T_0) + \tfrac{1}{2}\dot{f}_\mathrm{s}(T-T_0)^2\right], \tag{2}$$

where

$$T = t + \delta t = t + \frac{\vec{r} \cdot \vec{n}}{c} + \Delta T. \tag{3}$$

Here, T is the time of arrival of the signal at the solar system barycentre, ϕ_0 is the phase of the signal at a fiducial time T_0, $\vec{r}$ is the position of the detector with respect to the solar system barycentre, $\vec{n}$ is a unit vector in the direction of the neutron star, c is the speed of light and ΔT contains the relativistic corrections to the arrival time [18].

If f_s and $\dot{f}_s$ are known from (for example) radio observations, the signal can be *heterodyned* by multiplying the data by $\exp[-i\Psi(t)]$, low-pass filtered and resampled, so that the only time varying quantity remaining is the antenna pattern of the interferometer. We are left with a simple model with four unknown parameters h_0, ψ, ϕ_0 and ι. If there is an uncertainty in the frequency and frequency derivative then we have two additional parameters, the differences between the signal and heterodyne frequency and frequency derivatives, Δf and $\Delta \dot{f}$, giving a total of six unknown parameters.

A detailed description of the heterodyning procedure is presented elsewhere [7, 17]. Here we just provide a brief summary of this standard technique. The raw signal, $s(t)$, is centred near twice the rotation frequency of the neutron star, but is Doppler modulated due to the motion of the Earth and the orbit of the neutron star if it is in a binary system. The modulation bandwidth is typically 10^4 times less than the detector bandwidth, so one can greatly reduce the effective data rate by extracting this band and shifting it to zero frequency. In its standard form the result is one binned data point, B_k, every minute, containing all the relevant information from the original time series but at only 2×10^{-6} the original data rate. If the phase evolution has been correctly accounted for at this heterodyning stage, then the only time-varying component left in the signal will be the effect of the antenna pattern of the interferometer, as its geometry with respect to the neutron star varies with Earth rotation. Any small error, Δf, in the heterodyne frequency will cause the signal to oscillate at Δf (plus the residual Doppler shift).

The data points, B_k, are assumed to be uncorrelated. In this paper the B_k are generated using white Gaussian noise with $\mu = 0$ and $\sigma = 1$. The variance, σ_k, associated with each bin is therefore known be unity *a priori*, and the noise is uncorrelated between bins. For real data, this assumption may not hold. However, in practice we have found that when using sufficiently small bandwidths with GEO and LIGO data the noise is not significantly correlated between bins. It is also assumed that the noise is stationary over the 60 s of data contributing to each bin. This is also consistent with current instrumental performances.

3. The adaptive Metropolis–Hastings algorithm

After heterodyning, the signal on which we wish to carry out our MCMC analysis has the form [17]

$$y(t_k; \boldsymbol{a}) = \frac{1}{4} F_+(t_k; \psi) h_0 (1 + \cos^2 \iota)\, e^{i\Delta\Psi(t)} - \frac{i}{2} F_\times(t_k; \psi) h_0 \cos\iota\, e^{i\Delta\Psi(t)}, \tag{4}$$

where t_k is the time of the kth bin and $\boldsymbol{a} = (h_0, \psi, \phi_0, \cos\iota, \Delta f, \Delta \dot{f})$ is a vector of our unknown parameters. $\Delta\Psi(t)$ represents the residual phase evolution of the signal, equalling $\phi_0 + 2\pi[\Delta f (T - T_0) + \Delta \dot{f} (T - T_0)^2/2]$. The objective is to fit this model to the data

$$B_k = y(t_k; \boldsymbol{a}) + \epsilon_k, \tag{5}$$

where ϵ_k is assumed to be normally distributed noise with a mean of zero and known variance σ_k^2. Assuming statistical independence of the binned data points, B_k, the joint likelihood that these data $\boldsymbol{d} = \{B_k\}$ arise from a model with a certain parameter vector $\boldsymbol{a}$ is [17]

$$p(\boldsymbol{d}|\boldsymbol{a}) \propto \prod_k \exp\left[-\frac{1}{2}\left|\frac{B_k - y(t_k;\boldsymbol{a})}{\sigma_k}\right|^2\right] = \exp\left[\frac{-\chi^2(\boldsymbol{a})}{2}\right], \tag{6}$$

where

$$\chi^2(\boldsymbol{a}) = \sum_k \left|\frac{B_k - y(t_k;\boldsymbol{a})}{\sigma_k}\right|^2. \tag{7}$$

In order to draw any inference on the unknown parameter vector $\boldsymbol{a}$, we need the (posterior) probability of $\boldsymbol{a}$ given $\boldsymbol{d}$, which can be obtained from the likelihood via an application of Bayes' theorem. The unnormalized posterior density

$$p(\boldsymbol{a}|\boldsymbol{d}) \propto p(\boldsymbol{a})p(\boldsymbol{d}|\boldsymbol{a}) \tag{8}$$

is the product of the prior density of $\boldsymbol{a}$, $p(\boldsymbol{a})$, and the joint likelihood. Accordingly, appropriate priors have to be chosen for the particular parameters. In this study we use uniform priors with prior ranges $[-\pi, \pi]$, $[-\pi/4, \pi/4]$ and $[-1, 1]$ for the angle parameters ϕ_0, ψ and $\cos\iota$, respectively. For h_0 we also specify a uniform prior with boundary [0, 1000] in units of the rms noise [17]. For the frequency and spin-down uncertainty, we use suitable uniform priors with ranges of $\left[-\frac{1}{60}, \frac{1}{60}\right]$ Hz and $[-10^{-9}, 10^{-9}]$ Hz s^{-1} for Δf and $\Delta\dot{f}$, respectively.

The normalized posterior density $p(\boldsymbol{a}|\boldsymbol{d}) = p(\boldsymbol{a})p(\boldsymbol{d}|\boldsymbol{a})/p(\boldsymbol{d})$ cannot be evaluated analytically, so we use Monte Carlo methods to explore $p(\boldsymbol{a}|\boldsymbol{d})$. If we can simulate from $p(\boldsymbol{a}|\boldsymbol{d})$, we can estimate all interesting quantities, including the posterior means of all parameters from the corresponding sample means, to any desired accuracy by increasing the sample size.

However, drawing independent samples in a six-dimensional parameter space is not feasible. It has already been shown that MCMC methods can be used to parametrize gravitational wave signals of low signal-to-noise ratio [14] with four unknown parameters. These simulate a Markov chain, constructed so that its stationary distribution coincides with the posterior distribution and the sample path averages converge to the expectations. A minimal requirement for this is the irreducibility of the chain and hence the ability of the chain to reach all parts of the state space [11]. A specific MCMC technique is the MH algorithm [12, 13] which does not require the normalization constant, only the unnormalized posterior density of equation (8). We employed the MH algorithm for the four- and five-parameter pulsar detection problems [14]. The efficiency of the MH algorithm depends heavily on the choice of the proposal density. Intuition suggests that the closer the proposal distribution is to the target, the faster convergence to stationarity is achieved. Default choices such as a Gaussian proposal or a random walk result in very slow mixing for this six-parameter problem. To increase the speed of convergence, we employed an *adaptive* technique, adaptive in the sense that it allows the choice of proposal distribution to depend upon information gained from the already sampled states as well as the proposed but rejected states. The idea behind the delayed rejection algorithm specified by [15] is that persistent rejection, perhaps in particular parts of the state space, may indicate that locally the proposal distribution is badly calibrated to the target. Therefore, the MH algorithm is modified so that on rejection, a second attempt to move is made with a proposal distribution that depends on the previously rejected state. This adaptive Monte-Carlo method [15] was generalized for the variable dimension case [19] and renamed the 'delayed rejection method'. Since we have a fixed dimension problem here we implemented the original version [15], and also the generalization [19] that uses the reversible

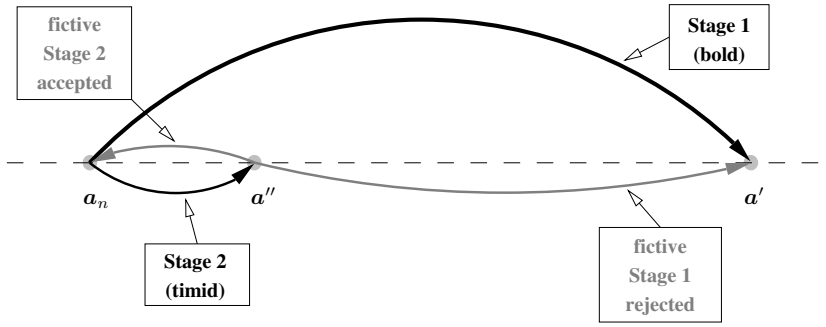

Figure 1. The delayed rejection method. In the case of rejection of the first, bold step a second, more timid move is proposed. In order to maintain the reversibility of the Markov chain the acceptance probability has to consider a fictive return path.

jump method. It turned out that the delayed rejection with the reversible jump method was not that beneficial for this particular problem and thus we will explain the original delayed rejection algorithm [15] here.

For the Metropolis–Hastings algorithm, a new state in a Markov chain is chosen first by sampling a candidate $\boldsymbol{a}'$ from a certain proposal distribution $q_1(\boldsymbol{a}'|\boldsymbol{a}_n)$ usually depending on the current state $\boldsymbol{a}_n$ and then accepting or rejecting it with a probability $\alpha_1(\boldsymbol{a}'|\boldsymbol{a}_n)$ depending on the distribution of interest. This rejection is essential for the convergence of the chain to the intended target distribution. The choice of a good proposal distribution is important to avoid persistent rejections in order to achieve good convergence of a chain. However in different parts of the state space different proposals are required. When a proposed MH move is rejected, a second candidate $\boldsymbol{a}''$ can be sampled with a different proposal distribution $q_2(\boldsymbol{a}''|\boldsymbol{a}', \boldsymbol{a}_n)$ that can depend on the previously rejected proposal. Since a rejection suggests a bad fit of the first proposal, a different form of proposal can be advantageous in the second stage. To preserve reversibility of the Markov chain and thus to comply with the detailed balance condition, the acceptance probabilities for both the first and the second stages are given by [20]

$$\alpha_1(\boldsymbol{a}'|\boldsymbol{a}_n) = \min\left(1, \frac{p(\boldsymbol{a}')p(\boldsymbol{d}|\boldsymbol{a}')q_1(\boldsymbol{a}_n|\boldsymbol{a}')}{p(\boldsymbol{a}_n)p(\boldsymbol{d}|\boldsymbol{a}_n)q_1(\boldsymbol{a}'|\boldsymbol{a}_n)}\right) \tag{9}$$

and

$$\alpha_2(\boldsymbol{a}''|\boldsymbol{a}_n) = \min\left(1, \frac{p(\boldsymbol{a}'')p(\boldsymbol{d}|\boldsymbol{a}'')q_1(\boldsymbol{a}'|\boldsymbol{a}'')q_2(\boldsymbol{a}_n|\boldsymbol{a}', \boldsymbol{a}'')[1-\alpha_1(\boldsymbol{a}'|\boldsymbol{a}'')]}{p(\boldsymbol{a}_n)p(\boldsymbol{d}|\boldsymbol{a}_n)q_1(\boldsymbol{a}'|\boldsymbol{a}_n)q_2(\boldsymbol{a}''|\boldsymbol{a}_n, \boldsymbol{a}')[1-\alpha_1(\boldsymbol{a}'|\boldsymbol{a}_n)]}\right), \tag{10}$$

respectively. Figure 1 illustrates the idea of delayed rejection. When the second stage proposal step is applied due to rejection of the first, the chain has, in order to preserve the reversibility, to imply a return path which comprises a fictive stationary Markov chain consisting of a fictive stage 1 proposal step from $\boldsymbol{a}''$ to $\boldsymbol{a}'$ which is rejected followed by an accepted fictive second stage move to $\boldsymbol{a}_n$ [19]. Although the delayed rejection method provides better acceptance rates over the two stages, cross-correlations between the parameters still impede convergence of the Markov chain. Preliminary runs reveal that especially the parameters Δf and $\Delta \dot{f}$, and to a certain extent h_0 and $\cos\iota$, are highly correlated after the Markov chain has found a potential mode. The consequence of which is poor mixing of the chain and therefore a reparametrization is required.

The coherence between Δf and $\Delta \dot{f}$ is obvious since the data are sampled from time t_{start} to t_{end}, where the heterodyned signal traverses a frequency from $f_{\text{start}} = \Delta f + \frac{1}{2}\Delta \dot{f} \cdot t_{\text{start}}$ to $f_{\text{end}} = \Delta f + \frac{1}{2}\Delta \dot{f} \cdot t_{\text{end}}$; time $t = 0$ is an epoch time when $f = \Delta f$. Hence it is much

more natural to work with these frequencies as new parameters and vary them with a certain correlation which influences $\Delta\dot{f}$ indirectly. The original parameters are then obtained by the simple linear transformation

$$\Delta f = f_{\text{start}} - \tfrac{1}{2}\Delta\dot{f} \cdot t_{\text{start}} \tag{11}$$

and

$$\Delta\dot{f} = 2 \cdot \frac{f_{\text{end}} - f_{\text{start}}}{t_{\text{end}} - t_{\text{start}}}. \tag{12}$$

Since the Jacobian of this transformation is constant the prior distributions for the new parameters f_{start} and f_{end} are flat as well.

Another cross-correlation can be observed between the parameters h_0 and $\cos\iota$ that arises from the fact that h_0 can be seen as a scaling factor and $\cos\iota$ as a nonlinear weighting between the plus and cross polarization parts of the model. As seen in equation (4), the plus part is multiplied by the factor $a_1 = \frac{1}{4}h_0(1 + \cos^2\iota)$ while the cross part encloses the term $a_2 = \frac{1}{2}h_0\cos\iota$. The original parameters can be derived from

$$h_0 = 2\left(a_1 + \sqrt{a_1^2 - a_2^2}\right), \tag{13}$$

and

$$\cos\iota = \frac{2a_2}{h_0}. \tag{14}$$

As mentioned above, the prior distribution of the parameters h_0 and $\cos\iota$ are chosen uniform with joint probability density function

$$f(h_0, \cos\iota) = \begin{cases} (2l_{h_0})^{-1}, & \text{if} \quad 0 \leqslant h_0 < l_{h_0}, \quad -1 \leqslant \cos\iota \leqslant 1, \\ 0, & \text{otherwise}, \end{cases} \tag{15}$$

where for this study $l_{h_0} = 1000$ in units of the rms noise. This implies a joint prior distribution for the parameters a_1 and a_2 of the form

$$g(a_1, a_2) = \left.\begin{cases} (2l_{h_0})^{-1}, & \text{if} \quad |a_2| \leqslant a_1 < \frac{4a_2^2 + l_{h_0}^2}{4l_{h_0}} \leqslant \frac{l_{h_0}}{2} \\ 0, & \text{otherwise} \end{cases}\right\} |\det J| \tag{16}$$

with Jacobian

$$\det J = \frac{2}{\sqrt{a_1^2 - a_2^2}}. \tag{17}$$

Since the Jacobian is positive for the above restrictions, we can write

$$g(a_1, a_2) = \begin{cases} \frac{1}{l_{h_0}\sqrt{a_1^2 - a_2^2}}, & \text{if} \quad |a_2| \leqslant a_1 < \frac{4a_2^2 + l_{h_0}^2}{4l_{h_0}} \leqslant \frac{l_{h_0}}{2}, \\ 0, & \text{otherwise}. \end{cases} \tag{18}$$

This joint prior density has the shape shown in figure 2. These reparametrizations result in a fast mixing Markov chain but still, the choice of a suitable proposal distribution is essential. Usually, a multivariate normal distribution is utilized for the proposal distributions $q_1(\boldsymbol{a}'|\boldsymbol{a}_n)$ and $q_2(\boldsymbol{a}''|\boldsymbol{a}', \boldsymbol{a}_n)$, with means equal to the current state and different variances depending on the stage. Larger variances are chosen for the 'bold' first stage steps, while smaller variances are more beneficial for the 'timid' second stage candidates. The covariance matrix has to comprise the correlation between the parameters f_{start} and f_{end} since this correlation indirectly controls the parameter $\Delta\dot{f}$ as mentioned above. Hence choosing proposals for f_{start} and f_{end} with a correlation of 1 would imply no change of $\Delta\dot{f}$ because both parameters are changed

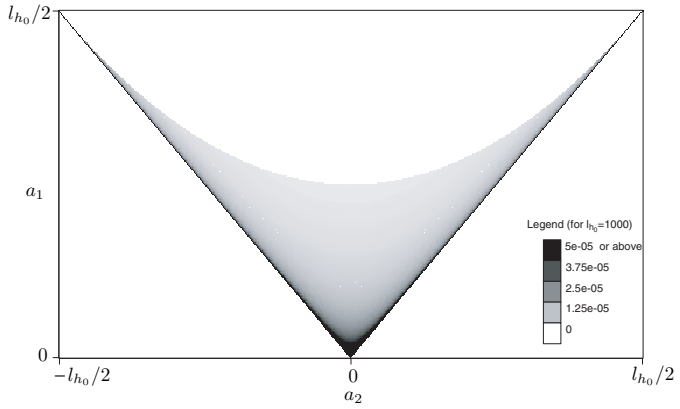

Figure 2. Joint prior density of a_1 and a_2 for a given boundary l_{h_0} for the parameter h_0.

in the same way, while a correlation of 0 would have a great impact on $\Delta\dot{f}$ since f_{start} and f_{end} are changed completely uncorrelated. Thus the correlation between f_{start} and f_{end} has to be treated randomly in order to control $\Delta\dot{f}$. Best results are obtained when a correlation of 0 is chosen with probability 0.5 for the bold moves of $\Delta\dot{f}$ and a correlation of almost 1 otherwise for timid moves of $\Delta\dot{f}$. The proposals for the parameters a_1 and a_2 are sampled independently since they represent scaling factors for the plus and cross polarization parts, respectively. Finally we have to consider the correlation between the original parameters ψ and ϕ_0 which are not reparametrized. Pilot runs show that they are highly correlated. Hence the proposal distribution is adapted accordingly.

Unfortunately, the posterior distribution features very narrow modes in a large parameter space that has to be scanned. Thus a simple normal distribution is not suitable for a proposal distribution as pilot runs have revealed. Instead, a proposal distribution with long tails and strong narrow mode is required. This can easily be achieved by generating a random sample between two boundaries b_l and b_h for the standard deviation of the proposal by generating a random weight for the weighted geometric mean of these two boundaries. Hence we sample standard deviations according to $\sigma = b_h^w b_l^{1-w}$, where $w \sim \beta(a, b)$ is beta distributed with parameters a and b. The resulting proposal distribution is symmetric with very long tails and a strong narrow mode. In order to obtain higher standard deviations for the first stage the choice of $w \sim \beta(2, 1)$ (with mean $\frac{2}{3}$) is adequate while for the second stage $w \sim \beta(1, 2)$ (with mean $1/3$) samples smaller standard deviations.

The implementation of the ideas outlined above leads to reasonable acceptance rates and hence to a much better convergence of the Markov chain. While during the burn-in period it is mainly the stage 1 candidates that are accepted, the Markov chain is driven mainly by stage 2 candidates after the burn-in. But still, the stationary distribution features many distinct modes that carry the risk of trapping the Markov chain. Therefore, we regard the posterior as a canonical distribution

$$\begin{aligned} p(\boldsymbol{a}|\boldsymbol{d}) &\propto p(\boldsymbol{a})p(\boldsymbol{d}|\boldsymbol{a}) \propto p(\boldsymbol{a})\exp\left[-\frac{\chi^2(\boldsymbol{a})}{2}\right] \\ &\propto \exp\left[-\frac{\chi^2(\boldsymbol{a}) - 2\log[p(\boldsymbol{a})]}{2}\right] \\ &\propto \exp[-\beta(\chi^2(\boldsymbol{a}) - 2\log[p(\boldsymbol{a})])] \end{aligned} \tag{19}$$

with inverse temperature β. During the burn-in period, this inverse temperature can pass through values starting at a low value (thus high temperature) and ending up at $\beta = \frac{1}{2}$ which coincides with the posterior distribution. This simulated annealing technique was introduced by Metropolis *et al* [12] and allows scanning of the whole parameter space by permitting larger steps. For the annealing schedule an exponential temperature curve is applied. For a certain number of iterations t_s, it starts with an inverse temperature β_0 until it reaches $\beta = \frac{1}{2}$. The inverse temperature follows the function

$$\beta(t) = \begin{cases} \beta_0 \exp\left[\frac{t}{t_s} \log\left(\frac{\beta}{\beta_0}\right)\right], & \text{if} \quad 0 \leqslant t \leqslant t_s, \\ \frac{1}{2}, & \text{if} \quad t > t_s, \end{cases} \tag{20}$$

depending on the current iteration t. Since the starting temperature is dependent on the data set which is influenced by the amplitude h_0 of the signal it has to be adapted accordingly.

4. Results with simulated signals

We have synthesized fictitious data, and passed it through our six-parameter MCMC routine. The presentation of results here is similar to that of the four- and five-parameter study of [14]. The artificial signals were embedded within white and normally distributed noise. The ability of the MCMC algorithm to successfully find the signal and estimate the six parameters was demonstrated, and is presented below. The artificial signals $s(t)$ were synthesized assuming a source at RA = 4 h 41 min 54 s and dec = $18^\circ 23' 32''$, as would be seen by the LIGO–Hanford interferometer. The signals were then added to noise; we assumed a signal at 300 Hz and a corresponding noise spectral density of that at frequency of $h(f) = 8 \times 10^{-23}$ Hz$^{-1/2}$. The amplitude of the signal used in our test runs was varied in the range from $h_0 = 4.0 \times 10^{-24}$ to 1.5×10^{-22}. The length of the data set corresponded to 14 400 samples or 10 days of data at a rate of one sample per minute (which was the rate used for the LIGO/GEO S1 analysis described in [7]).

In figure 3 we display the MCMC generated posterior probability distribution functions (pdfs) for an example signal. The *real* parameters for this signal were $h_0 = 1.5 \times 10^{-22}$, $\psi = 0.4$, $\phi_0 = 1.0$ (both in radians), $\cos\iota = 0.878$, $\Delta f = 7.0 \times 10^{-3}$ Hz and $\Delta\dot{f} = -2.5 \times 10^{-10}$ Hz s^{-1}. For this example, the programme ran for 10^6 iterations. For a signal this large only about 2.5×10^4 iterations were needed for the burn-in, and these data are discarded from the analysis. Short-term correlations in the chain were eliminated by 'thinning' the remaining terms; we kept every 250th item in the chain. The bandwidth of the Gaussian kernel density estimator was chosen according to Silverman [21] as 0.9 times the minimum of the standard deviation and the interquartile range divided by 1.34 times the sample size to the negative one-fifth power.

In this example the MCMC yielded median values, 95% posterior probability intervals and MCMC standard errors (se) of $h_0 = 14.91 \times 10^{-23}$ (13.41×10^{-23} to 15.84×10^{-23}), se$(h_0) = 1.686 \times 10^{-2}$, $\psi = 0.439$ (-0.552 to 0.707), se$(\psi) = 7.505 \times 10^{-3}$, $\phi_0 = 0.964$ (0.696 to 1.958), se$(\phi) = 7.510 \times 10^{-3}$, $\cos\iota = 0.884$ (0.828 to 0.988), se$(\cos\iota) = 1.103 \times 10^{-3}$, $\Delta f = 6.999\,997\,72 \times 10^{-3}$ Hz ($6.999\,992\,17 \times 10^{-3}$ Hz to $7.000\,003\,14 \times 10^{-3}$ Hz), se$(\Delta f) = 2.386 \times 10^{-10}$ and $\Delta\dot{f} = -2.499\,9541 \times 10^{-10}$ Hz s^{-1} ($-2.500\,0767 \times 10^{-10}$ Hz s^{-1} to $-2.499\,8272 \times 10^{-10}$ Hz s^{-1}), se$(\Delta\dot{f}) = 2.386 \times 10^{-10}$. The 95% posterior probability interval is specified by the 2.5% and 97.5% percentiles of $p(a_i|\boldsymbol{d})$. The MCMC standard error gives a measure of how much the sample mean, as a point estimate of the true posterior mean, changes over repeated MCMC simulations. This precision depends on the number of iterations and the degree of autocorrelation within the sample. We used Geweke's [22] method, often referred to as 'time-series standard error' which is based on

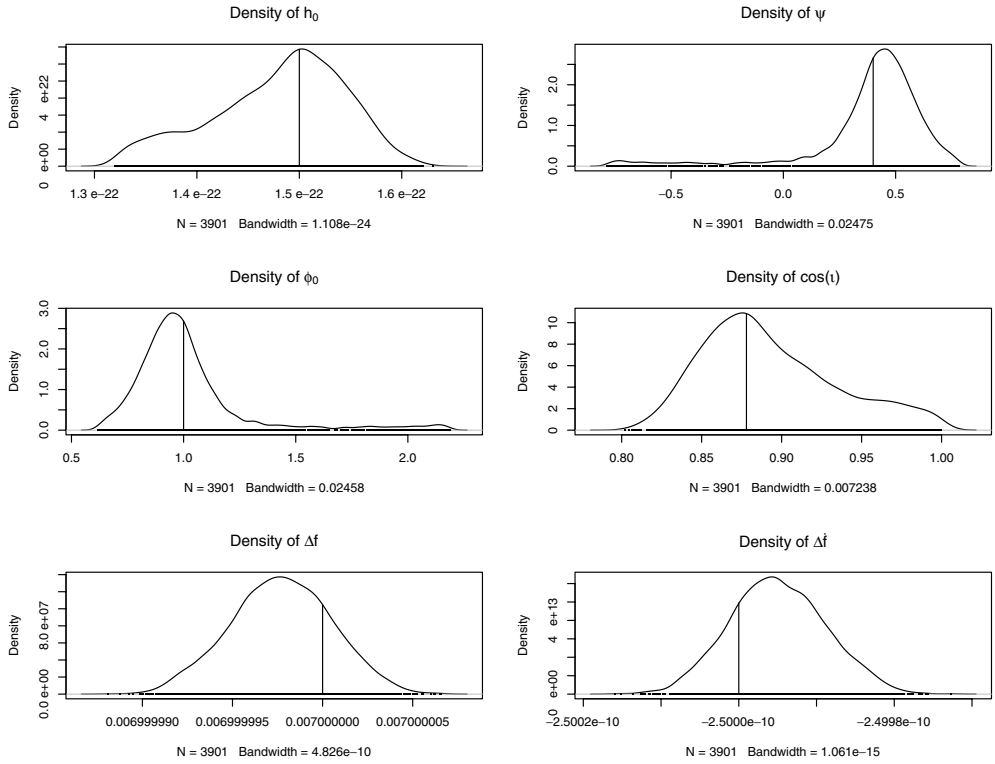

Figure 3. MCMC estimates of the posterior pdf (kernel density) for the six parameters h_0, ψ, ϕ_0, $\cos\iota$, Δf and $\Delta \dot{f}$. This synthesized signal had *real* parameters of $h_0 = 15\times10^{-23}$, $\psi = 0.4$, $\phi_0 = 1.0$, $\cos\iota = 0.878$, $\Delta f = 7.0 \times 10^{-3}$ Hz and $\Delta \dot{f} = -2.5 \times 10^{-10}$ Hz s^{-1}. The mean of the h_0 distribution here is 14.84×10^{-23}. The vertical lines show the *real* parameter values.

estimating the spectral density. The interested reader should note the paper by Geyer [23] who develops improved window-estimates for the MCMC standard error by calculating the 'optimal' bandwidth using specific properties of the autocovariances of a Markov chain.

With the noise level used, $h(f) = 8 \times 10^{-23}$, we were able to successfully detect signals with amplitudes of $h_0 \geqslant 4.0 \times 10^{-24}$ with 10 days of data. This should be compared with the results presented in [14] where with just four parameters (h_0, ψ, ϕ_0 and $\cos\iota$), we were able to confidently detect signals with an amplitude four times smaller. The addition of the new frequency parameters has the disadvantage of complicating the search due to the corresponding increase in the size of the parameter space. For our study, we let the initial burn-in of the Markov chain last for as long as 3.5×10^5 iterations, and if the signal was not found by this time the search was terminated. It may be possible to find smaller signals with a longer burn-in.

Figure 4 shows the MCMC estimated posterior for the smallest value of the parameter h_0 that we were able to identify with the MCMC code. The true parameter values for this run were $h_0 = 4.0\times10^{-24}$, $\psi = 0.4$, $\phi_0 = 1.0$, $\iota = 0.5$ ($\cos\iota = 0.878$), $\Delta f = 7.0\times10^{-3}$ Hz and $\Delta \dot{f} = -2.5 \times 10^{-10}$ Hz s^{-1}. In this run, the MCMC yielded a mean value and 95% posterior probability interval of $h_0 = 4.8 \times 10^{-24}$ (3.4×10^{-24} to 7.4×10^{-24}), se(h_0) $= 2.164 \times 10^{-3}$. Figure 5 displays the MCMC estimated posterior for the parameters Δf and $\Delta \dot{f}$, which provides mean values and 95% posterior probability intervals of $\Delta f = 7.0 \times 10^{-3}$ Hz

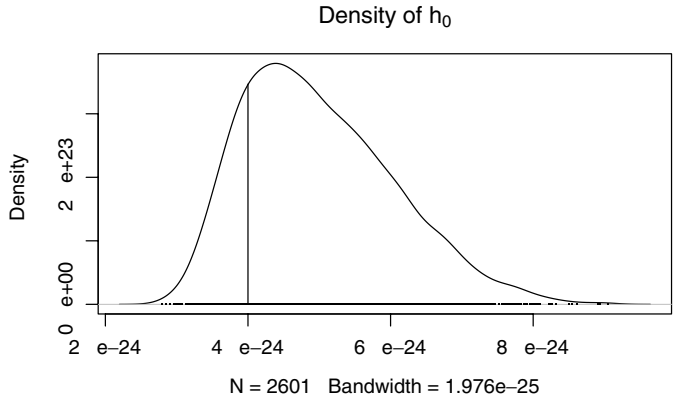

Figure 4. MCMC estimate of the posterior pdf (kernel density) for the parameter h_0 from a six-parameter search using synthesized data. The real parameter value for this signal was $h_0 = 4.0 \times 10^{-24}$ (vertical line). This was the smallest signal detectable by the MCMC method for the noise level used.

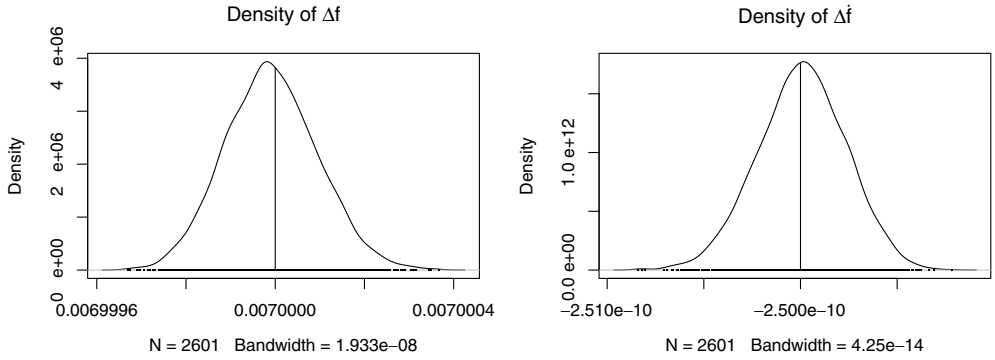

Figure 5. MCMC estimate of the posterior pdfs (kernel densities) for the parameters Δf and $\Delta \dot{f}$ from a six-parameter search using synthesized data with the smallest detectable signal $h_0 = 4.0 \times 10^{-24}$. The real parameters for this signal (vertical lines) were $\Delta f = 7.0 \times 10^{-3}$ Hz and $\Delta \dot{f} = -2.5 \times 10^{-10}$ Hz s^{-1}.

(6.9998×10^{-3} Hz to 7.0002×10^{-3} Hz), se$(\Delta f) = 3.460 \times 10^{-9}$ and $\Delta \dot{f} = -2.500 \times 10^{-10}$ Hz s^{-1} (-2.505×10^{-10} Hz s^{-1} to -2.496×10^{-10} Hz s^{-1}), se$(\Delta \dot{f}) = 2.922 \times 10^{-10}$. As can be seen from figures 4 and 5, even with small signal level it is still possible to extract the most astrophysically important parameters. For this MCMC run there were a total of 10^6 iterations, with the first 3.5×10^5 as the burn-in. For low amplitude signals that were successfully found by the MCMC routine good estimates for the parameters h_0, $\cos \iota$, Δf and $\Delta \dot{f}$ could be made; however, the posterior pdfs for the angular parameters ψ and ϕ_0 tended to spread uniformly across the allowed range of their *a priori* distributions.

5. Discussion and conclusions

In the simplest application, the method demonstrated here could complement searches for signals from known pulsars [7, 17]; our method could be used to verify the frequency and frequency derivative values. The real advantage of the technique would come about in a

search for a signal at a known location, but where the frequency information pertaining to the neutron star is not well known; a search for a signal from SN1987A [10] would be a possible application. In the demonstration here the heterodyning process provides a band of 1/60 Hz. It would be straightforward to expand this search to a bandwidth of 5 Hz by running the code on 300 processors, a task easily accomplished on a cluster of computers. For 10 days of data, it takes a single 2.8 GHz personal computer approximately 1 h to conduct about 3.3×10^4 iterations of our MCMC code. There are more iterations done per time interval at the beginning of a run because at that time more stage-1 steps are accepted.

The code has also been successfully tested on real interferometer data, where the noise spectral density is coloured, containing artificially injected signals; these results will appear in a subsequent publication. The heterodyning process and the noise estimation procedure [17] were still successful. The MCMC routine was also successful and robust with these data. In summary, we believe that these MCMC methods offer great potential benefits for gravitational radiation searches where the signals depend on a large number of parameters.

Acknowledgments

This work was supported by the National Science Foundation grants PHY-0071327 and PHY-0244357, The Royal Society of New Zealand Marsden Fund Grant UOA 204, the University of Auckland Research Committee, the Natural Sciences and Engineering Research Council of Canada, Universities UK and the University of Glasgow.

References

[1] Cutler C 2002 *Phys. Rev.* D **66** 084025
[2] Bildsten L 1998 *Astrophys. J.* **501** L89
[3] Willke B *et al* 2002 *Class. Quantum Grav.* **19** 1377
[4] Abramovici A *et al* 1992 *Science* **256** 325
[5] Caron B *et al* 1996 *Nucl. Phys. Suppl.* **48** 107
[6] Tsubono K 1997 *Gravitational Wave Detection* ed K Tsubono, M-K Fujimoto and K Kurodo (Tokyo: Universal Academic) pp 183-91
[7] Abbott B *et al* 2003 *Phys. Rev.* D **69** 082004
[8] Jaranowski P, Królak A and Schutz B F 1998 *Phys. Rev.* D **58** 063001
[9] Brady P and Creighton T 2000 *Phys. Rev.* D **61** 082001
[10] Middleditch J *et al* 2000 *New Astron.* **5** 243
[11] Gilks W R, Richardson S and Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall)
[12] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[13] Hastings W K 1970 *Biometrika* **57** 97
[14] Christensen N, Dupuis R J, Woan G and Meyer R 2004 *Preprint* gr-qc/0402038 (*Phys. Rev* D at press)
[15] Tierney L and Mira A 1999 *Stat. Med.* **18** 2507
[16] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **4598** 671
[17] Dupuis R J and Woan G 2003 A Bayesian method to search for periodic gravitational waves *Preprint*
[18] Taylor J H 2002 *Phys. Rev.* D **66** 084025
[19] Green P J and Mira A 2001 *Biometrika* **88** 1035
[20] Mira A 1998 Ordering, slicing and splitting Monte Carlo Markov chain *PhD thesis* University of Minnesota
[21] Silverman B W 1986 *Density Estimation for Statistics and Data Analysis* (London: Chapman and Hall)
[22] Geweke J 1992 *Bayesian Statistics 4* ed J M Bernardo, J O Berger, A P Dawid and A F M Smith (Oxford: Oxford University Press) pp 169–93
[23] Geyer C J 1992 *Stat. Sci.* **7** 473