The University of Auckland

Digital thesis consent form

Author of thesis	Richard Umstaetter
Title of thesis	Bayesian Strategies for Gravitational Radiation Data Analysis
Name of degree	PhD

I agree that the University of Auckland Library may make a digital copy of my thesis available for consultation for the purposes of research and private study. It is understood that any person wishing to publish or otherwise use material from my thesis will require my permission before any material from the thesis is published or used.

This consent is in addition to any permissions I have already given for supply to the collection of another prescribed library, or for supply to any person in accordance with the provisions of Section 56 of the Copyright Act 1994.

I certify that the digital copy deposited with the University is the same as the final officially approved version of my thesis. Except in the circumstances set out below no emendation of content has occurred and if there are any minor variations in formatting they are the result of the conversion to digital format.

[X] I confirm that my thesis does not contain material, the copyright for which belongs to a third party

or

[ ] I confirm that I have obtained written permission to use all third party copyright material in my thesis and attach copies of each of the permissions required.

or

[ ] I confirm that where third party copyright material was included in my thesis and I have not been able to obtain written permission to include this material in the digital version of the thesis I have:

a) removed the third party copyright material from the thesis; and
b) fully referenced the deleted materials; and
c) where possible, provided links to electronic sources of the material

(Please tick the box which applies.)

Signed: ...Richard Umstaetter... (author of thesis)

Date: ...8.1.2007...

Bayesian Strategies for Gravitational Radiation Data Analysis

Richard Umstätter

A thesis
submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy in Statistics,

The University of Auckland, 2006

Abstract

This work addresses the exploration of Bayesian MCMC methods applied to problems in gravitational wave physics. The thesis consists of two parts. In the first part a Bayesian Markov chain Monte Carlo technique is presented 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 Metropolis-Coupled Markov Chain Monte Carlo. Results will be given for different synthesized data sets in order to demonstrate the algorithm's behaviour for different observation lengths and signal-to-noise ratios. The probability of detecting weak signals is assessed by a model comparison, based on the BIC, between a model that postulates a signal and one that postulates solely noise within the data.

The second part of the thesis adresses the tremendous data analysis challenges for the Laser Interferometer Space Antenna (LISA) with the need to account for a large number of gravitational wave signals from compact binary systems expected to be present in the data. The basis of a Bayesian method is introduced that can address this challenge, and its effectiveness is demonstrated on a simplified problem involving one hundred synthetic sinusoidal signals in noise. The reversible jump Markov chain Monte Carlo technique is deployed to infer simultaneously the number of signals present, the parameters of each identified signal, and the noise level. This approach is specifically focused on the detection of a large number of sinusoids with separation of sinusoids that are close in frequency. A robust post-processing technique handles the label switching problem by a frequency interval separation technique with a subsequent classification according to a mixed model approximation. The algorithm therefore tackles the detection and parameter estimation problems simultaneously, without the need to evaluate formal model selection criteria, such as the Akaike Information Criterion or explicit Bayes factors. The method produces results which compare very favorably with classical spectral techniques.

Acknowledgements

A key feature of modern science is its synergy between traditional disciplines. I would therefore like to thank first of all my supervisors Dr. Renate Meyer and Dr. Nelson Christensen who brought this exciting interdisciplinary project into being and for guiding me through the different worlds of Bayesian statistics and gravitational wave physics. I would also like to extend my thanks to the research group at the University of Glasgow and in particular Dr. Graham Woan for his very helpful inputs.

Special thanks go to my colleagues at the Department of Statistics and in particular to Christian Röver and Dr. Andreas Berg for many fertile discussions.

This work was supported by the Royal Society of New Zealand Marsden Fund Grant UOA204.

Contents

1 Introduction **1**

2 Parameter estimation for GW-signals **5**

2.1 Introduction 5

2.2 The gravitational wave signal 7

2.3 The Bayesian full probability model 9

2.4 The adaptive Metropolis-Hastings algorithm 11

2.4.1 The delayed rejection method 11

2.4.2 Re-parameterisation 13

2.4.3 The choice of proposal distributions 16

2.4.4 The Metropolis-Coupled MCMC 19

2.5 Results with simulated signals 24

2.5.1 Simulation results for strong signals 24

2.5.2 Simulation results for weak signals 31

2.6 The detection of weak signals 38

2.6.1 The problem with assessing the signal detection from MCMC outputs 38

2.6.2 Derivation of a theoretical detection probability 42

2.6.3 Signal detection results for different scenarios 51

2.7 Discussion 58

2.8 Outlook 60

3 Bayesian estimation of confusion noise **61**

3.1 Introduction 61

3.2 The Bayesian full probability model 65

3.3 Sampling from the posterior distribution 68
3.3.1 The RJMCMC for model determination 68
3.3.2 The delayed rejection method for parameter estimation 72
3.3.3 Updating the noise parameter 75
3.3.4 Updating the hyperparameter g^2 75
3.3.5 Initial values 76
3.4 Identifying the sinusoids 76
3.4.1 Identifiability constraints 77
3.4.2 Relabelling algorithms 78
3.4.3 Interval separation of sinusoids by their frequency using a label invariant loss function 82
3.4.4 Further classification using a mixture model approximation 90
3.5 Simulation results 94
3.6 Discussion 118

4 Conclusions **121**

List of Figures

2.1 Diagram of the delayed rejection method. 13
2.2 Joint prior density of a_1 and a_2 for a given boundary l_{h_0} for the parameter h_0. 16
2.3 Log-posterior values for all Metropolis-Coupled chains. 25
2.4 MCMC estimates of the posterior pdf (kernel density) for the six parameters of a pulsar. 27
2.5 MCMC estimates of the posterior pdf (kernel density) for the six parameters of a pulsar. 28
2.6 Joint 2-D posterior distributions of different parameter pairs. . 30
2.7 MCMC estimates of the posterior pdf (kernel density) for the six parameters of a pulsar. 34
2.8 Log-posterior values for all Metropolis-Coupled chain. 36
2.9 Log-posterior values for all Metropolis-Coupled chain. 36
2.10 Trace plots of the four parameters of a pulsar which are of main interest. 37
2.11 Expected model probabilities for a signal within the data. . . . 52
2.12 Expected model probabilities for a signal within the data marginalised over the inclination and polarisation angle. 54
2.13 Expected model probabilities for a signal within the data of different length marginalised over the inclination and polarisation angle. 55
2.14 Expected model probabilities for a signal within the data of different length marginalised over the sky location, inclination and polarisation angle. 56

2.15 Comparison of three expected model probabilities for a signal within the data of different length marginalised over the sky location, inclination and polarisation angle. 57
2.16 Doppler modulation of the heterodyne frequency in case of an offset between reference sky location for heterodyning and actual sky location of the pulsar. 59

3.1 Example for the label switching problem with optimal chosen intervals to separate the sinusoidal components. 84
3.2 The two steps of preallocating the sinusoidal components. . . . 89
3.3 Signal-to-noise ratios (SNRs) for each individual sinusoid used in the artificial data set, ordered by SNR. 95
3.4 Posterior model probabilities and corresponding noise levels. . 96
3.5 Marginal MCMC posterior distribution for different parameter pairs of the blendoid. 100
3.6 Quantile-Quantile-plot of the marginal posterior distribution of the frequency of sinusoids. 102
3.7 Marginal MCMC posterior distribution of a single, isolated sinusoid and of a pair of sinusoids that are close in frequency. . 103
3.8 Gradually magnified area of a sinusoid pair with frequency separation 0.00047 and one that has a very low signal-to-noise ratio. 105
3.9 Comparison of true spectral lines, Bayesian spectral density estimate, and classical Schuster periodogram. 107
3.10 Gradually magnified area of a group of four sinusoids. 108
3.11 Comparison of true spectral lines, Bayesian spectral density estimate, and classical Schuster periodogram. 109
3.12 Gradually magnified area of a sinusoid pair with frequency separation 0.00044. 110
3.13 Comparison of true spectral lines, Bayesian spectral density estimate, and classical Schuster periodogram. 111

3.14 Model probabilities of a pair of sinusoids in Gaussian noise with varying difference in frequency and uniform prior on the amplitudes. 116
3.15 Model probabilities of a pair of sinusoids in Gaussian noise with varying difference in frequency and uniform g-prior on the amplitudes. 117

List of Tables

2.1 Median values obtained by MCMC, 95% posterior probability intervals (p.c.i.) and MCMC standard errors for the data set of length 14 400 bins. 29

2.2 Median values obtained by MCMC, 95% posterior probability intervals and MCMC standard errors for the data set of length 60 000 bins. 29

2.3 Ratios for all six parameters of the 95% posterior probability interval ranges between Tab. 2.1 and Tab. 2.2. 31

2.4 MCMC yielded median values, 95% posterior probability intervals and MCMC standard errors for the data set of length 14 400 bins. 35

3.1 Frequency intervals that maximise the number of single occupancies in model $\mathcal{M}_{99}$ and percentages of individual occupancy numbers. 96

3.2 Posterior means and 95% posterior credibility intervals of the frequency and the Cartesian amplitude. 112