Inference on inspiral signals using LISA MLDC data

Christian Röver[1], Alexander Stroeer[2,3], Ed Bloomer[4], Nelson Christensen[5], James Clark[4], Martin Hendry[4], Chris Messenger[4], Renate Meyer[1], Matt Pitkin[4], Jennifer Toher[4], Richard Umstätter[1], Alberto Vecchio[2,3], John Veitch[4] and Graham Woan[4]

[1] Department of Statistics, The University of Auckland, Auckland, New Zealand
[2] School of Physics & Astronomy, University of Birmingham, Birmingham, UK
[3] Department of Physics & Astronomy, Northwestern University, Evanston, IL, USA
[4] Department of Physics & Astronomy, University of Glasgow, Glasgow, UK
[5] Physics & Astronomy, Carleton College, Northfield, MN, USA

Abstract. In this paper we describe a Bayesian inference framework for analysis of data obtained by LISA. We set up a model for binary inspiral signals as defined for the Mock LISA Data Challenge 1.2 (MLDC), and implemented a Markov chain Monte Carlo (MCMC) algorithm to facilitate exploration and integration of the posterior distribution over the 9-dimensional parameter space. Here we present intermediate results showing how, using this method, information about the 9 parameters can be extracted from the data.

PACS numbers: 04.80.Nn, 02.70.Uu.

Submitted to: *Classical and Quantum Gravity*

1. Introduction

Once the LISA gravitational wave observatory is launched and operational, it is certain to measure a vast number of signals from a wide range of sources. Because the data will contain a superposition of individually modulated signals blended with noise, sophisticated methods will be required to disentangle individual signals and consistently infer their parameters. Bayesian inference provides a means to approach such complex problems, allowing one to quantify the information that is buried in the data in a coherent manner [1, 2, 3]. We are convinced that these techniques will be useful, if not essential, to analyse the data that will be obtained through LISA. Bayesian procedures, in conjunction with Markov chain Monte Carlo (MCMC) methods, have successfully been applied for the analysis of ground-based GW measurements [4, 5, 6], as well as in the context of LISA [7, 8], and in particular in the presence of source confusion [9]. Related work on MCMC methods for LISA inspiral analysis can also be found in [10, 11].

The authors have gathered as the '*Global LISA Inference Group*' (GLIG) and have set out to implement such an analysis framework for LISA data. We have developed some generic code modules providing vital components that can be adapted to allow analysis given different model specifications. In response to the first round of the Mock LISA Data Challenges (MLDC) [12], we present the results of an analysis targeted at binary inspiral signals as addressed in MLDC Challenge 1.2. Within the same context, we approached MLDC Challenge 1.1, containing white dwarf binary systems. These results are presented in [13].

We implemented an MCMC sampler to perform the integration of the posterior probability distribution of the 9 parameters that determine the waveform of a binary inspiral GW signal, and present results illustrating the parameter information that can be extracted from the data. Due to the tight schedule we did not manage to enhance the MCMC sampler's convergence capabilities sufficiently to get results for the 'blind search' data as well. For now we present results for the 'training' data only.

2. Inference framework

2.1. The Bayesian approach

We use a Bayesian framework to perform inference on gravitational wave signals observed by LISA, aiming for the *information* about parameters that can be derived from the data. Information about parameters here is formulated in terms of probability distributions over the parameter space. First the prior knowledge about parameters ϑ needs to be properly specified in the *prior distribution* $p(\vartheta)$. Then parameters and data y are linked by defining the *likelihood* $p(y|\vartheta)$ that describes how the observables come about for given parameter values. Inference eventually is done via the parameters' *posterior distribution* $p(\vartheta|y)$, which expresses the information about the parameter values *conditional on the oberved data*. The posterior distribution is given by $p(\vartheta|y) \propto p(\vartheta)\, p(y|\vartheta)$, as a consequence from *Bayes' theorem* [1, 2, 3]. Inference on the measured signal's parameters (or other properties) requires integration of the posterior distribution over the parameter space, since one is usually interested in determining figures like posterior expectations, marginal (posterior) densities, or confidence regions. We approach the problem using Monte Carlo integration, for which we implemented a Markov chain Monte Carlo (MCMC) algorithm. The algorithm eventually is supposed to be able to reliably *find* the global mode(s) in the posterior distribution and then perform the *integration*, i.e. sample from the posterior [3, 14].

2.2. Data and parameters

A gravitational wave (GW) signal is measured by LISA by monitoring the changes in proper distance between the three satellites as they are orbiting the Sun. The data is sampled every 15 seconds, which is also about the time it takes for a photon to travel from one satellite to another. The measured response is not a simple '1:1' mapping of the signal waveform to the data, especially when the signal wavelength is of the order or below LISA's armlength. Moreover, as LISA orbits, the response will also be modulated by Doppler effects and be affected by the change in the baseline orientations over time.

The data produced by the spacecraft trio is combined to form three time-delay-interferometry (TDI) variables, X, Y and Z [15]. These can be linearly recombined into three stochastically independent components, out of which two are sensitive to GW signals (A and E) and one component is only noise (T) [16]. In the following we will only be concerned with the former two variables, A and E.

In the restricted 2.0 PN approximation, the 9 parameters defining a binary inspiral's GW signal measured by LISA are chirp mass (m_c), mass ratio (η), coalescence phase (ϕ_c), coalescence time (t_c), ecliptic latitude (θ), ecliptic longitude (φ), luminosity distance (D), polarisation (ψ) and inclination angle (ι) [17].

2.3. Model

For given data y from a single TDI variable the likelihood function is defined as a function of the parameters ϑ by

$$p(y|\vartheta) \;\propto\; \exp\left(-\sum_f \frac{|\tilde{y}(f)-\tilde{s}(f,\vartheta)|^2}{S_n(f)}\right), \tag{1}$$

where $\tilde{y}$ and $\tilde{s}$ are the (numerical) discrete Fourier transforms of data and signal waveform respectively, and S_n is the variable's (one-sided) noise spectral density [18]. The data (y) going into the likelihood here are the 'A' and 'E' TDI variables [16]. Assuming that these are stochastically independent, the likelihood then is the product of the individual variables' likelihoods. The signal waveform $\tilde{s}$ to which the data are matched is the corresponding TDI response to the GW signal implied by the parameters ϑ. The noise spectrum S_n refers to the noise in the corresponding (A and E) TDI variables.

3. Implementation

In order to infer the measured signal's parameters, one first needs to *find* the global mode(s) in the posterior distribution, and then also to '*explore*' the mode(s), i.e. simulate posterior samples. We tackle that problem using MCMC methods, an approach that in particular requires many likelihood evaluations. Likelihood computations are computationally expensive, since they require several time-consuming steps: Given a parameter set ϑ, one first needs to compute the $+/\times$ polarisation waveforms emitted by the inspiral event, then the TDI response of the LISA interferometer to the GW signal, and finally its Fourier transform, before parameters can be related to the data through the likelihood. Since most of these (and more) steps are common between a wide range of different types of analyses, we set up our software in a modular style so that parts would be reusable and shareable in form of modules. See our accompanying paper [13] for another application that shares parts of the same code. So far, this also includes a common framework to store and manipulate data internally, an interface to the *lisaXML* data format [17], and the availability of Fourier transformation and spectrum estimation capabilities based on the *FFTW* library [19]. Most importantly, the derivation of LISA's response (in terms of X/Y/Z or A/E TDI variables) to a gravitational wave signal (given in terms of $+/\times$ polarisations, direction of source, and polarisation angle) was needed. Here we resorted to the *LISA Simulator* [20, 21], that was also used for the generation of the MLDC data, and which is coded in C, allowing us to easily incorporate it into our code and stay consistent with the provided data. Originally, the LISA Simulator was not intended to do its computations repeatedly and quickly, and it was possible to speed up the code by storing and reusing some intermediate steps. We implemented a

Table 1. Computation speed of the MCMC code for different amounts of data (on an Intel Xeon 2.4 GHz processor). Most of the computation time (more than 95%) goes into deriving A/E TDI responses from the $+/\times$ GW waveforms.

amount of data		seconds
days	samples	per iteration
364	2^{21}	146
182	2^{20}	75
91	2^{19}	38
46	2^{18}	19
23	2^{17}	10

simple Metropolis-algorithm [3, 14] to do inference on binary inspiral signals as defined for MLDC challenge 1.2.1 [17]. The eventual computation speed, depending on how much data are processed, is shown in table 1. A similar implementation has proven successful in the context of ground-based GW measurements [6], and we are currently working on tuning and extending the basic algorithm.

4. Results

We applied the above framework to the data in MLDC challenge 1.2.1. The signal waveform was generated following the description given in [17], and we estimated the A/E variables' noise spectral densities based on the section of data where the signal was absent.

We ran the code on the 'training' data set, starting from the true parameter values, and, due to the low computation speed, only considered the last 2^{17} samples (corresponding to 23 days of measurements) before coalescence for the analysis. The resulting speed of the MCMC sampler still was rather slow, producing a posterior sample every 10 seconds. By only considering the last part of of the signal before coalescence we are of course neglecting some information, but since the SNR of the injected signal was very high (almost 500), and most of that is actualised in the last phase immediately before coalescence, we will still be left with a high SNR. On the other hand, we will especially lose information about the location parameters (θ, φ), since these are encoded in the long-term evolution of the signal, so we might find an increased degeneracy between these two parameters. As the efficiency of our code continues to improve we will of course be analysing larger sections of data.

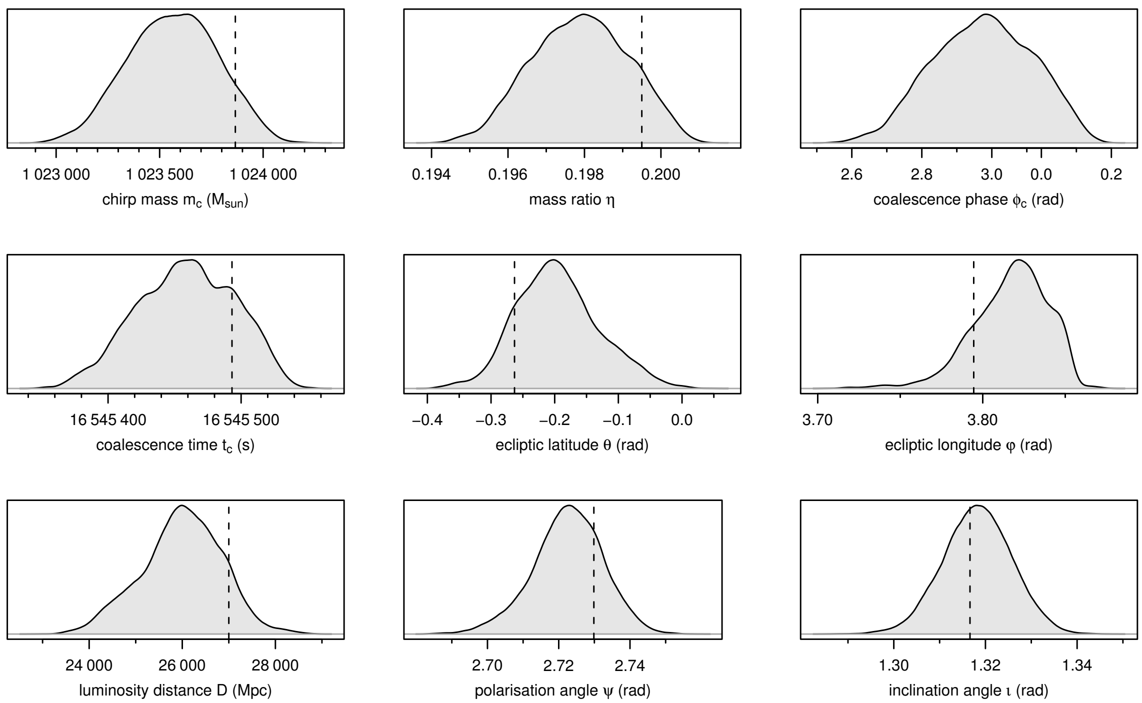

Figure 1. Marginal posterior densities for all 9 parameters. Dashed lines indicate the true values.

Figure 1 shows the marginal posterior distributions for all 9 individual parameters in comparison to the true parameter values (there is no true value shown for ϕ_c, since we are using a different parametrisation: *coalescence phase* instead of *initial phase*). The true parameter values are all well covered by the posterior distribution, not only in these 1-dimensional projections, but also for all bivariate distributions, two examples of which are shown in figure 2; this demonstrates the consistency of the applied inference framework. Table 2 shows some summary statistics characterizing the posterior distribution, and relating it to the true parameter values. As one can see from figure 2, there is much posterior correlation, or degeneracy, between the parameters. In particular, there are two groups of parameters that are highly correlated with each other: firstly the two mass parameters, coalescence time and phase (m_c, η, t_c, ϕ_c), and secondly, the two sky location parameters and the luminosity distance (θ, φ, D). Correlation coefficients of parameter pairs within these groups are as high as 0.90–0.99, which greatly complicates sampling from the posterior.

We also ran the code on the 'blind' challenge 1.2.1 data set, for which we did not know the true parameter values, but due to the code's speed and the size of the parameter space it would not converge and produce results in time before the submission

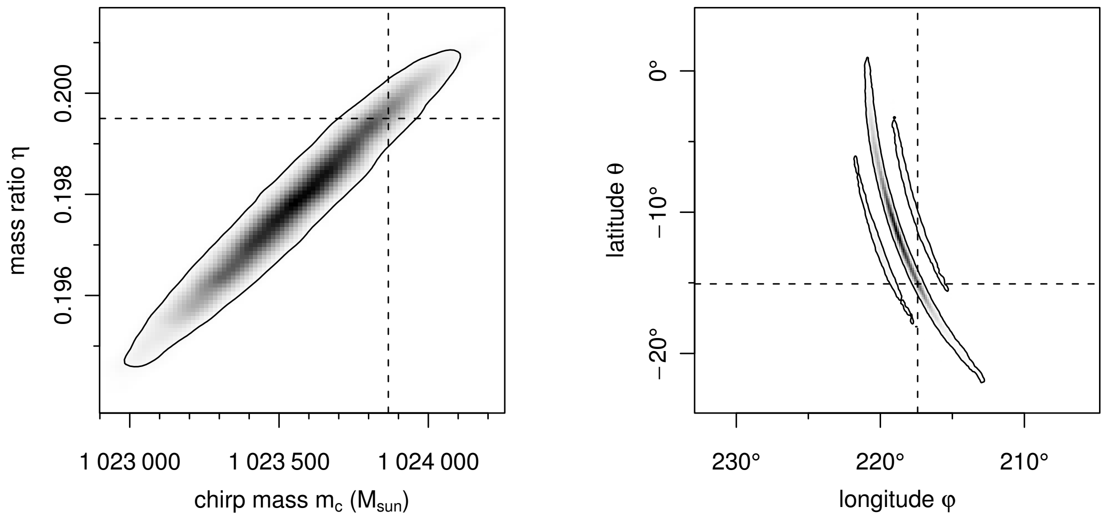

Figure 2. Marginal joint posterior densities and 99% credibility regions for two pairs of parameters. Dashed lines indicate the true values.

Table 2. Some summary statistics characterizing the marginal posterior distributions for individual parameters. Mean and standard deviation describe location and accuracy, and the 99% central credibility intervals contain the corresponding parameter with 99% probability, given the data at hand.

	mean	st.dev.	99% c.c.i.	true	unit
chirp mass m_c	1 023 564	215	(1 023 033, 1 024 054)	1 023 866	$M_\odot$
mass ratio η	0.1979	0.0013	(0.1948, 0.2006)	0.1995	
coalescence phase ϕ_c	2.97	0.14	(2.63, 0.13)		rad
coalescence time t_c	16 545 459	36	(16 545 370, 16 545 535)	16 545 493	s
ecliptic latitude θ	−0.195	0.065	(−0.357, −0.019)	−0.263	rad
ecliptic longitude φ	3.817	0.023	(3.736, 3.859)	3.795	rad
luminosity distance D	25 991	864	(23 754, 28 195)	27 000	Mpc
polarisation ψ	2.7226	0.0099	(2.6943, 2.7464)	2.7299	rad
inclination angle ι	1.3182	0.0075	(1.2978, 1.3368)	1.3166	rad

deadline for MLDC round 1.

While the MCMC algorithm is working in principle and performing the posterior integration, more tuning is necessary to enhance its optimisation properties, i.e. its capabilities of finding modes by itself, and its efficiency in manoeuvering through parameter space. Better convergence properties are crucial not only to enhance the algorithm's overall applicability, but also to make sure not to be missing further posterior modes that may be of relevance.

5. Conclusions

We have presented a Bayesian inference framework for analysis of GW signals as measured by LISA. We ran a basic MCMC algorithm on data simulating a binary inspiral measurement from the first round of the Mock LISA Data Challenges (MLDC). In a related effort [13], sharing parts of the same code, we applied a similar model to the analysis of signals from white dwarf binary systems. The MCMC implementation so far is a simple Metropolis algorithm, and the results illustrate that this approach ultimately allows one to extract and express the information about signal parameters contained in the data in a coherent manner. While the *integration* of the posterior distribution over the parameter space is fully functional, more work needs to be done on the MCMC sampler's *optimisation* capabilities as well as its efficiency. We are working on a preprocessing stage to the MCMC algorithm to provide rough parameter estimates as starting values for the MCMC sampler. We are also currently extending the Metropolis-sampler to a *parallel tempering algorithm* [14, 6] in a parallel implementation [22]. The underlying model will also need to be generalised by including the noise spectrum as an unknown, which might just mean the introduction of an additional 'Gibbs step' in the MCMC sampler [3, 14].

Acknowledgments

This work was supported by the Marsden Fund Council from Government funding administered by the Royal Society of New Zealand, the National Science Foundation, the Fulbright Scholar Program and the Packard Foundation.

References

[1] E. T. Jaynes. *Probability theory: The logic of science*. Cambridge University Press, Cambridge, 2003.

[2] P. C. Gregory. *Bayesian logical data analysis for the physical sciences*. Cambridge University Press, Cambridge, 2005.

[3] A. Gelman, J. B. Carlin, H. Stern, and D. B. Rubin. *Bayesian data analysis*. Chapman & Hall / CRC, Boca Raton, 1997.

[4] R. J. Dupuis and G. Woan. Bayesian estimation of pulsar parameters from gravitational wave data. *Physical Review D*, 72(10):1022002, November 2005.

[5] C. Röver, R. Meyer, and N. Christensen. Bayesian inference on compact binary inspiral gravitational radiation signals in interferometric data. *Classical and Quantum Gravity*, 23(15):4895–4906, August 2006.

[6] C. Röver, R. Meyer, and N. Christensen. Coherent Bayesian inference on compact binary inspirals using a network of interferometric gravitational wave detectors. *Physical Review D*, 75(6):062004, March 2007.

[7] E. D. L. Wickham, A. Stroeer, and A. Vecchio. A Markov chain Monte Carlo approach to the study of massive black hole binary systems with LISA. *Classical and Quantum Gravity*, 23(19):S819–S827, September 2006.

[8] A. Stroeer, J. Gair, and A. Vecchio. Automatic Bayesian inference for LISA data analysis strategies. *Arxiv preprint gr-qc/0609010*, September 2006.

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

[10] N. J. Cornish and E. K. Porter. MCMC exploration of supermassive black hole binary inspirals. *Classical and Quantum Gravity*, 23(19):S761–S767, October 2006.

[11] N. J. Cornish and E. K. Porter. The search for supermassive black hole binaries with LISA. *Arxiv preprint gr-qc/0612091*, December 2006.

[12] K. A. Arnaud et al. An overview of the Mock LISA Data Challenges. In S. M. Merkowitz and J. C. Livas, editors, *Laser Interferometer Space Antenna: 6th International LISA Symposium*, volume 873 of *AIP conference proceedings*, pages 619–624, November 2006.

[13] A. Stroeer, J. Veitch, C. Röver, E. Bloomer, N. Christensen, J. Clark, M. Hendry, C. Messenger, R. Meyer, M. Pitkin, J. Toher, R. Umstätter, A. Vecchio, and G. Woan. Inference on white dwarf binary systems using the first round Mock LISA Data Challenges data sets. *Arxiv preprint 0704.0048 [gr-qc]*, March 2007.

[14] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall / CRC, Boca Raton, 1996.

[15] J. W. Armstrong, F. B. Estabrook, and M. Tinto. Time-delay interferometry for space-based gravitational wave searches. *The Astrophysical Journal*, 527(2):814–826, December 1999.

[16] T. A. Prince, M. Tinto, S. L. Larson, and J. W. Armstrong. LISA optimal sensitivity. *Physical Review D*, 66(12):122002, December 2002.

[17] MLDC taskforce. Document for challenge 1, draft v1.0, August 2006. URL `http://svn.sourceforge.net/viewvc/*checkout*/lisatools/Docs/challenge1.pdf`.

[18] L. S. Finn and D. F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Physical Review D*, 47(6):2198–2219, March 1993.

[19] M. Frigo and S. G. Johnson. FFTW 3.0.1, a C subroutine library for computing the discrete Fourier Transform (DFT), 2003. URL `http://www.fftw.org`.

[20] N. J. Cornish, L.J. Rubbo, and O. Poujade. The LISA Simulator, version 2.1.1, June 2006. URL `http://www.physics.montana.edu/LISA`.

[21] J. Cornish, N and L. J. Rubbo. LISA response function. *Physical Review D*, 67(2):022001, January 2003.

[22] M. Paprzycki and P. Stpiczyński. A brief introduction to parallel computing. In E. J. Kontoghiorghes, editor, *Handbook of parallel computing and statistics*, chapter 1, pages 3–41. Chapman & Hall / CRC, Boca Raton, 2006.