

Inference on inspiral signals using LISA MLDC data

Christian Röver¹, Alexander Stroeer^{2,3}, Ed Bloomer⁴,
Nelson Christensen⁵, James Clark⁴, Martin Hendry⁴, Chris Messenger⁴,
Renate Meyer¹, Matt Pitkin⁴, Jennifer Toher⁴, Richard Umstätter¹,
Alberto Vecchio^{2,3}, John Veitch⁴ and Graham Woan⁴

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 Department of Physics & Astronomy, Carleton College, Northfield, MN, USA

Received 31 March 2007

Published 19 September 2007

Online at stacks.iop.org/CQG/24/S521

Abstract

In this paper, we describe a Bayesian inference framework for the 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 nine-dimensional parameter space. Here, we present intermediate results showing how, using this method, information about the nine parameters can be extracted from the data.

PACS numbers: 04.80.Nn, 02.70.Uu

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 for approaching such complex problems, allowing one to quantify the information that is buried in the data in a coherent manner [1–3]. We are convinced that these techniques will be useful, if not essential, for analysing 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 gravitational wave (GW) measurements [4–6], 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 [21, 22].

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) [10], 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 [11].

We implemented an MCMC sampler to perform the integration of the posterior probability distribution of the nine 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 is eventually done via the parameters' *posterior distribution* $p(\vartheta|y)$, which expresses the information about the parameter values *conditional on the observed data*. The posterior distribution is given by $p(\vartheta|y) \propto p(\vartheta)p(y|\vartheta)$, as a consequence from *Bayes' theorem* [1–3]. 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 such as posterior expectations, marginal (posterior) densities or confidence regions. We approach the problem using Monte Carlo integration, for which we implemented an MCMC algorithm. The algorithm is eventually 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, 12].

2.2. Data and parameters

A 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 are sampled every 15 s, 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 arm length. Moreover, as LISA orbits, the response will also be modulated by Doppler effects and affected by the change in the baseline orientations over time.

The data produced by the spacecraft trio are combined to form three time-delay-interferometry (TDI) variables, X , Y and Z [13]. 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) [14]. In the following, we will only be concerned with the former two variables, A and E. In the restricted 2.0 PN approximation, the nine 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), polarization (ψ) and inclination angle (i) [15].

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

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

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 [16]. The data (y) going into the likelihood here are the 'A' and 'E' TDI variables [14]. 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 $+/\times$ polarization 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 the form of modules. See our accompanying paper [11] 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 [15], and the availability of Fourier transformation and spectrum estimation capabilities based on the *FTW* library [17]. 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 $+/\times$ polarizations, direction of source and polarization angle) was needed. Here, we resorted to the *LISA Simulator* [18, 19], 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 simple Metropolis algorithm [3, 12] to do inference on binary inspiral signals as defined for MLDC Challenge 1.2.1 [15]. The eventual computation speed, depending on how much data are processed, is shown in table 1. A similar implementation has been shown to be successful in the context of ground-based GW measurements [6], and we are currently working on tuning and extending the basic algorithm.

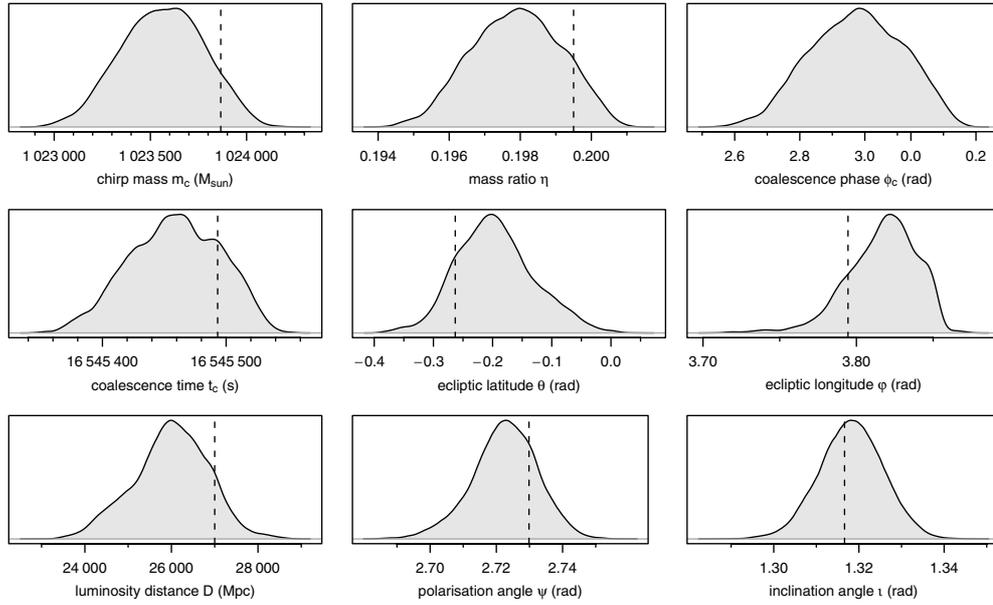


Figure 1. Marginal posterior densities for all nine parameters. Dashed lines indicate the true values.

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 per iteration
Days	Samples	
364	2^{21}	146
182	2^{20}	75
91	2^{19}	38
46	2^{18}	19
23	2^{17}	10

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 [15], 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 s. By only considering the last part 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 actualized 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

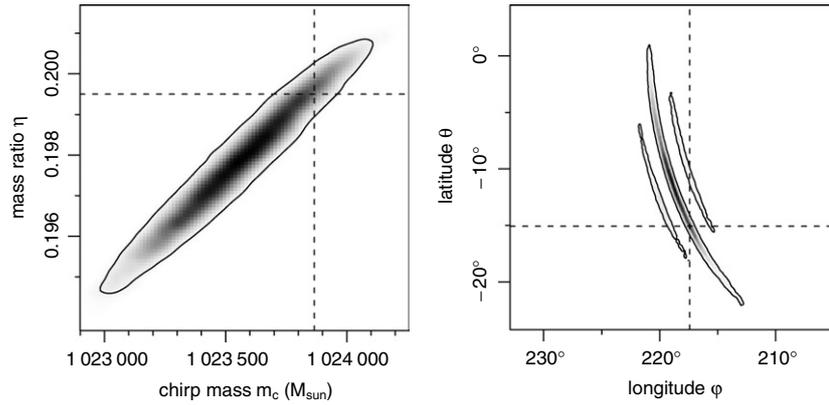


Figure 2. Marginal joint posterior densities and 99% credibility regions for two pairs of parameters. Dashed lines indicate the true values.

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.

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	Standard deviation	99% c.c.i.	True	Unit
Chirp mass m_c	1023 564	215	(1023 033, 1024 054)	1023 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
Polarization 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

Figure 1 shows the marginal posterior distributions for all nine individual parameters in comparison to the true parameter values (there is no true value shown for ϕ_c , since we are using a different parametrization: *coalescence phase* instead of *initial phase*). The true parameter values are all well covered by the posterior distribution, not only in these one-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 for the submission 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 optimization properties, i.e. its capabilities of finding modes by itself and its efficiency in manoeuvring through parameter space. Better convergence properties are crucial not only to enhance the algorithm's overall applicability, but also to make sure it does not miss further posterior modes that may be of relevance.

5. Conclusions

We have presented a Bayesian inference framework for the 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 [11], 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 *optimization* 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* [6, 12] in a parallel implementation [20]. The underlying model will also need to be generalized by including the noise spectrum as an unknown, which might just mean the introduction of an additional 'Gibbs step' in the MCMC sampler [3, 12].

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] Jaynes E T 2003 *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press)
- [2] Gregory P C 2005 *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge: Cambridge University Press)
- [3] Gelman A, Carlin J B, Stern H and Rubin D B 1997 *Bayesian Data Analysis* (Boca Raton, FL: Chapman and Hall/CRC Press)
- [4] Dupuis R J and Woan G 2005 Bayesian estimation of pulsar parameters from gravitational wave data *Phys. Rev. D* **72** 1022002
- [5] Röver C, Meyer R and Christensen N 2006 Bayesian inference on compact binary inspiral gravitational radiation signals in interferometric data *Class. Quantum Grav.* **23** 4895–906
- [6] Röver C, Meyer R and Christensen N 2007 Coherent Bayesian inference on compact binary inspirals using a network of interferometric gravitational wave detectors *Phys. Rev. D* **75** 062004
- [7] Wickham E D L, Stroeer A and Vecchio A 2006 A Markov chain Monte Carlo approach to the study of massive black hole binary systems with LISA *Class. Quantum Grav.* **23** S819–27
- [8] Stroeer A, Gair J and Vecchio A 2006 Automatic Bayesian inference for LISA data analysis strategies *Preprint gr-qc/0609010*
- [9] Umstätter R, Christensen N, Hendry M, Meyer R, Simha V, Veitch J, Vigeland S and Woan G 2005 Bayesian modeling of source confusion in LISA data *Phys. Rev. D* **72** 022001

- [10] Arnaud K A *et al* 2006 An overview of the Mock LISA Data Challenges. *Laser Interferometer Space Antenna: 6th Int. LISA Symposium* ed S M Merkowitz and J C Livas *AIP Conf. Proc.* **873** 619–24
- [11] Stroerer A *et al* 2007 Inference on white dwarf binary systems using the first round Mock LISA Data Challenges data sets *Class. Quantum Grav.* **24** S541
- [12] Gilks W R, Richardson S and Spiegelhalter D J 1996 *Markov chain Monte Carlo in Practice* (Boca Raton, FL: Chapman and Hall/CRC Press)
- [13] Armstrong J W, Estabrook F B and Tinto M 1999 Time-delay interferometry for space-based gravitational wave searches *Astrophys. J.* **527** 814–26
- [14] Prince T A, Tinto M, Larson S L and Armstrong J W 2002 LISA optimal sensitivity *Phys. Rev. D* **66** 122002
- [15] MLDC Taskforce 2006 *Document for Challenge 1* Draft v1.0 (August) http://svn.sourceforge.net/viewvc/*checkout*/lisatools/Docs/challenge1.%pdf
- [16] Finn L S and Chernoff D F 1993 Observing binary inspiral in gravitational radiation: one interferometer *Phys. Rev. D* **47** 2198–219
- [17] Frigo M and Johnson S G 2003 FFTW 3.0.1: a C subroutine library for computing the discrete Fourier transform (DFT) <http://www.fftw.org>
- [18] Cornish N J, Rubbo L J and Poujade O 2006 *The LISA Simulator* version 2.1.1 <http://www.physics.montana.edu/LISA>
- [19] Cornish J N and Rubbo L J 2003 LISA response function *Phys. Rev. D* **67** 022001
- [20] Paprzycki M and Stpiczynski P 2006 A brief introduction to parallel computing *Handbook of Parallel Computing and Statistics* ed E J Kontoghiorghes (Boca Raton, FL: Chapman and Hall/CRC Press) chapter 1 pp 3–41
- [21] Cornish N J and Porter E K 2006 MCMC exploration of supermassive black hole binary inspirals *Class. Quantum Grav.* **23** S761–7
- [22] Cornish N J and Porter E K 2006 The search for supermassive black hole binaries with LISA *Preprint gr-qc/0612091*