

OPEN ACCESS

Nested sampling as a tool for LISA data analysis

To cite this article: Jonathan R Gair *et al* 2010 *J. Phys.: Conf. Ser.* **228** 012010

View the [article online](#) for updates and enhancements.

You may also like

- [THE HUNT FOR EXOMOONS WITH KEPLER \(HEK\). I. DESCRIPTION OF A NEW OBSERVATIONAL PROJECT](#)
D. M. Kipping, G. Á. Bakos, L. Buchhave et al.
- [EXONEST: BAYESIAN MODEL SELECTION APPLIED TO THE DETECTION AND CHARACTERIZATION OF EXOPLANETS VIA PHOTOMETRIC VARIATIONS](#)
Ben Placek, Kevin H. Knuth and Daniel Angerhausen
- [Aurora: A Generalized Retrieval Framework for Exoplanetary Transmission Spectra](#)
Luis Welbanks and Nikku Madhusudhan



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Nested Sampling as a tool for LISA data analysis

**Jonathan R Gair¹, Farhan Feroz², Stanislav Babak³, Philip Graff²,
Michael P Hobson², Antoine Petiteau³ and Edward K Porter⁴**

¹ Institute of Astronomy, Madingley Road, CB3 0HA, Cambridge, UK

² Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK

³ Max-Planck-Institut für Gravitationsphysik, Am Mühlenberg 1, 14476, Potsdam, Germany

⁴ APC, UMR 7164, Université Paris 7 Denis Diderot, 10, rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France

E-mail: jgair@ast.cam.ac.uk

Abstract. Nested sampling is a technique for efficiently computing the probability of a data set under a particular hypothesis, also called the Bayesian Evidence or Marginal Likelihood, and for evaluating the posterior. MULTINEST is a multi-modal nested sampling algorithm which has been designed to efficiently explore and characterize posterior probability surfaces containing multiple secondary solutions. We have applied the MULTINEST algorithm to a number of problems in gravitational wave data analysis. In this article, we describe the algorithm and present results for several applications of the algorithm to analysis of mock LISA data. We summarise recently published results for a test case in which we searched for two non-spinning black hole binary merger signals in simulated LISA data. We also describe results obtained with MULTINEST in the most recent round of the Mock LISA Data Challenge (MLDC), in which the algorithm was used to search for and characterise both spinning supermassive black hole binary inspirals and bursts from cosmic string cusps. In all these applications, the algorithm found the correct number of signals and efficiently recovered the posterior probability distribution. Moreover, in most cases the waveform corresponding to the best a-posteriori parameters had an overlap in excess of 99% with the true signal.

1. Introduction

The data stream from the proposed space-based gravitational wave detector, the Laser Interferometer Space Antenna (LISA) [1], is expected to contain thousands of overlapping gravitational wave signals from many different types of source, including galactic compact binaries, the inspirals and mergers of supermassive black hole binaries and extreme-mass-ratio inspirals. Disentangling and correctly identifying these many sources is a challenging problem for data analysis. Research in this area has been stimulated by a sequence of Mock LISA Data Challenges (MLDCs), of which there have been four to date [2, 3, 4, 5, 6]. The most recent of these, round 3, finished at the end of April 2009. Several of the LISA sources have degeneracies in their parameter spaces, which lead to the existence of multiple modes in the posterior probability surface, some of which can be nearly equal in probability to the true solution. In the early rounds of the MLDC, this caused problems for searches for extreme-mass-ratio inspirals in particular, as Markov Chain Monte Carlo algorithms tended to get stuck on secondary modes and did not find the primary [5, 6, 7]. Data sets containing multiple sources of a given type will also have highly multi-modal posterior probabilities when explored with single-source waveform models. To date, multiple-source extraction has mostly been tackled by iterative source identification

and subtraction, but simultaneous identification of modes is likely to be computationally more efficient.

Nested sampling [8] is a technique for efficiently computing the Bayesian evidence of a data set under a particular model hypothesis, but also returns the posterior over the model parameters as a by-product. The MULTINEST algorithm [9] is an efficient implementation of nested sampling, which is designed to deal with posteriors containing multiple modes of arbitrary shapes. It has proven to be effective in dealing with many varied problems in astronomy and particle physics [10, 11, 12, 13]. Moreover, the algorithm works without using any information about the type of signal space being explored, other than in the computation of the likelihood at individual points in the parameter space. As the posterior probability surfaces for LISA sources are expected to be highly multi-modal, MULTINEST is well-suited to tackling data analysis for LISA as well. In this article we describe the application of MULTINEST to characterization of various simulated LISA data sets, both in a controlled study [14] and for the analysis of two types of source in the last round of the MLDC. Nested sampling has previously been used in a gravitational wave context for model selection for ground based interferometers [15], and was a key element in a hybrid evolutionary algorithm that also proved effective for exploring the multi-modal posteriors of LISA data sets [16]. However, the work described here represents the first use of MULTINEST in a gravitational wave data analysis context.

The article is organised as follows. In section 2 we describe the MULTINEST algorithm and in section 3 we describe results for our controlled study and for the analysis of two of the MLDC round 3 data sets. We finish with a discussion of future applications of the algorithm in section 4.

2. The MULTINEST Algorithm

The Bayesian evidence associated with a hypothesis, $\mathcal{H}$, for data, $\mathbf{D}$, depending on parameters, Θ , is the factor that appears in the denominator of Bayes theorem

$$P(\Theta|\mathbf{D}, \mathcal{H}) = \frac{P(\mathbf{D}|\Theta, \mathcal{H})P(\Theta|\mathcal{H})}{P(\mathbf{D}|\mathcal{H})}$$

$$P(\mathbf{D}|\mathcal{H}) \equiv \mathcal{Z} \equiv \int P(\mathbf{D}|\Theta, \mathcal{H})P(\Theta, \mathcal{H}) d^N \Theta \quad (1)$$

The evidence, $\mathcal{Z}$, can be used to choose between alternative hypotheses, but it is time-consuming to evaluate as it is a multi-dimensional integral. Nested sampling is a Monte Carlo method for evaluating evidences efficiently [8], which works by replacing the multi-dimensional integral (1) with a one dimensional integral over the prior volume, X ,

$$X(\lambda) = \int_{\mathcal{L}(\Theta) > \lambda} P(\Theta|\mathcal{H}) d^N \Theta \quad \Rightarrow \quad \mathcal{Z} = \int_0^1 \mathcal{L}(X) dX \quad (2)$$

Here, $\mathcal{L}(\Theta) \equiv P(\mathbf{D}|\Theta, \mathcal{H})$ is the likelihood. If the likelihood is evaluated on a nested sequence of contours of decreasing prior volume, the integral (2) can be computed using, e.g., the trapezium rule. This is illustrated in Figure 1.

This algorithm is implemented in practice using a set of N live points initially drawn from the prior. At each iteration i the point of lowest likelihood, with likelihood $\mathcal{L}_i$, is removed from the set and replaced by a point of higher likelihood, $\mathcal{L} > \mathcal{L}_i$, sampled from the prior. At each iteration the prior volume is thus reduced by a factor t drawn from $P(t) = Nt^{N-1}$. The algorithm travels through nested shells of likelihood as the prior volume reduces.

The challenge is to efficiently sample new points from the prior with likelihood $\mathcal{L} > \mathcal{L}_i$. The MULTINEST algorithm [9] achieves this by partitioning the current set of live points into a set of (possibly overlapping) ellipsoids. An ellipsoid is then chosen at random and a new point sampled uniformly from within it. This is repeated until a point with likelihood $\mathcal{L} > \mathcal{L}_i$ is found. This

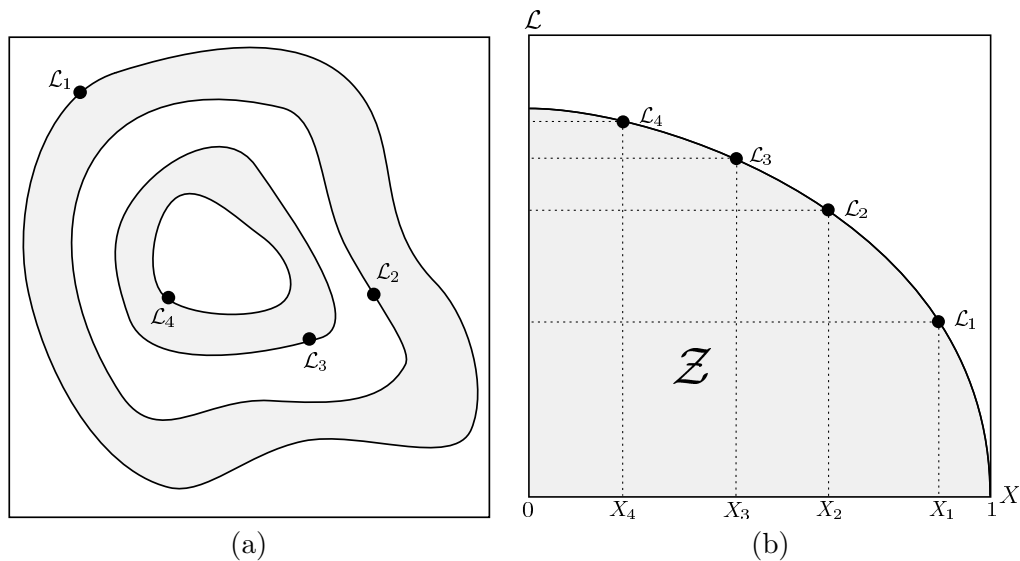


Figure 1. Cartoon illustrating (a) the posterior of a two dimensional problem; and (b) the transformed $\mathcal{L}(X)$ where the prior volumes, X_i , are associated with each likelihood $\mathcal{L}_i$.

ellipsoidal decomposition allows MULTINEST to identify and separate different modes in the posterior, making it an ideal algorithm for exploring multi-modal posterior probability surfaces. Although designed for evidence computation, the algorithm also returns the posterior and it is for characterising multi-modal posteriors that we have used the algorithm so far. We note that MULTINEST is not problem-specific in the sense that the particular model being used only enters in the likelihood evaluation and not in how the algorithm updates the live point set.

3. Results

3.1. Non-spinning black holes

As a first application of MULTINEST to gravitational wave data analysis, we generated a simulated LISA data set containing two signals from non-spinning supermassive black hole mergers [14]. One of the sources coalesced during the observation, while the second coalesced shortly after the observation ended. Due to the nature of the LISA response, there is a degeneracy between positions on the sky that are antipodal to one another. We therefore expected the posterior to contain at least four modes. We searched for these signals in two stages — the first stage involved searching only for the five intrinsic parameters of the source (two masses, time of coalescence and sky position) and used the generalised $\mathcal{F}$ -statistic [17, 18] to maximize automatically over the four extrinsic parameters (inclination, distance, polarization and phase at coalescence); the second stage involved searching the full nine dimensional parameter space, but with tight priors on the intrinsic parameters based on the results of the first stage.

The search took ~ 2.5 hours on a single 3GHz processor and successfully identified 11 posterior modes, of which four were the two true and two antipodal modes that we expected to see. The parameters for both sources were recovered to very high precision, well within the expected error as estimated from the Fisher Matrix. Full results and further details can be found in [14].

3.2. Mock LISA Data Challenge Round 3

The third round of the MLDC consisted of five data sets. The MULTINEST algorithm was used successfully to search two of these — challenge 3.2 which was a 2 year data stream containing 4–6 signals from inspiralling spinning supermassive black hole binaries; and challenge 3.4 which

Source	3.2.1	3.2.2	3.2.3	3.2.4	3.2.5
Overlap A	0.992	0.999	0.999	0.961	0.449
Overlap E	0.992	0.999	0.999	0.967	0.410

Table 1. Overlaps between the true waveforms and the maximum a-posteriori solutions identified by MULTINEST for each of the five spinning black hole binary sources in the MLDC challenge 3.2 data set. Overlaps are given for both the *A* and *E* TDI data channels.

was a 1 month data set containing an unknown number (drawn from a Poisson distribution with mean 5) of bursts from cosmic string cusps.

3.2.1. Challenge 3.2: spinning supermassive black hole binary inspirals For this search, we divided the data set into sections according to estimates of the plunge times for the sources as computed from a time-frequency analysis of the data. These sections were then searched with MULTINEST to obtain parameter estimates. We identified and recovered parameters for 5 sources, which was the correct number. The fifth source, the weakest, was only found after subtracting the best estimate for the fourth source from the data stream. The overlaps of the recovered sources with the true signals are tabulated in Table 1. Sources 3.2.2–3.2.4 were correctly recovered, although the dominant mode identified for 3.2.2 and 3.2.4 was the mode antipodal on the sky to the true solution. Source 3.2.1 was recovered with high overlap, but given this source was the brightest in the data set, with $\text{SNR} \sim 2000$, a 0.8% mismatch leaves a significant residual. This arose in part because of the approximations to the detector response model that we used, and subsequent runs after the MLDC deadline using the full detector response have been much closer to the true parameters. Source 3.2.5, which was a low-SNR source that did not coalesce within the observation window, was not correctly identified at the time of the MLDC deadline, but this was due to a shortage of time. This source has since been recovered correctly. More details on the search and results will be presented elsewhere.

3.2.2. Challenge 3.4: cosmic string bursts For this search, we divided the 1 month release data into segments of 8192s in length and these were searched with MULTINEST. We also used segments of length 16384s and 32768s to cross-check the results. We included the LISA response in our waveform model of the cosmic string burst using the static LISA approximation [19]. We identified three sources and this was again the true number that were present in the data set. The overlaps of the solutions identified by MULTINEST with the true waveforms were in excess of 99.7% for both the *A* and *E* channels in all three cases. This is sufficient to ensure that the residual after subtraction of the sources would be smaller than the instrumental noise in the detector. The recovered maximum a-posteriori parameters did not match the true values very closely, however. This was a consequence of the large degeneracies in the cosmic string parameter space, which arise because for burst sources the direction to the source cannot be well determined. This is illustrated in Figure 2, which shows the 2D and 1D marginalised posteriors for each of the six parameters that describe the cosmic string burst. We show these posteriors, as recovered by MULTINEST, for the second of the three burst signals present in the MLDC challenge data release. Degeneracies can be clearly seen in the posteriors for sky position and time of coalescence, which arise from symmetries in the response of the detector at low frequency. These features were also seen and discussed in [20]. Further details of our methods and results can be found in a separate paper [21].

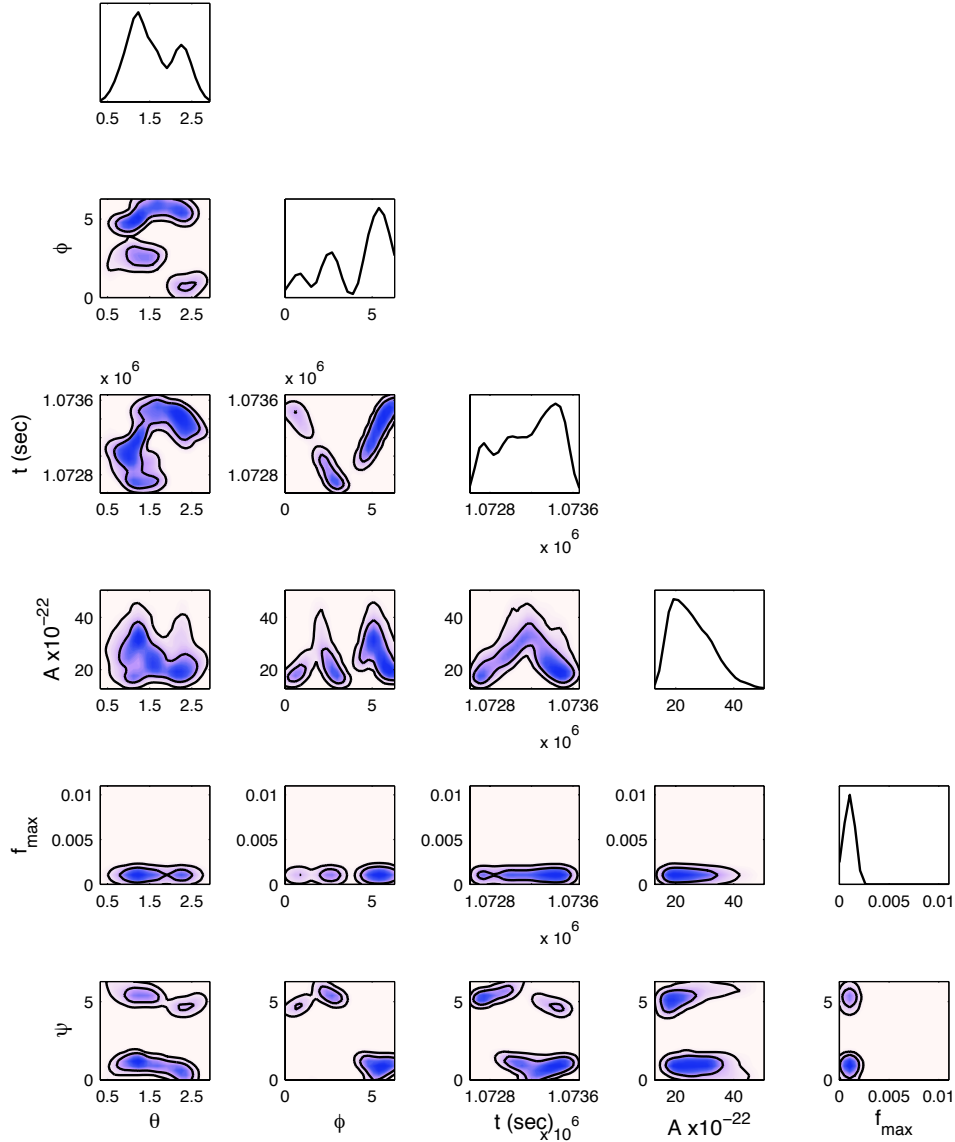


Figure 2. Two-dimensional marginalised posteriors as recovered by MULTINEST in the search for the second of the cosmic string bursts in the MLDC challenge 3.4 data set. The parameters, from top-to-bottom and left-to-right, are colatitude, longitude, burst time, burst amplitude, burst break frequency and waveform polarization. At the top of each column we also show the one-dimensional posterior for the column parameter.

4. Discussion

We have described the application of MULTINEST, a multi-modal nested sampling algorithm, to various problems in gravitational wave data analysis for LISA. Nested sampling is a method for computing evidences and posteriors efficiently, and MULTINEST is an implementation of nested sampling which is optimized for analysing posterior probability surfaces containing multiple modes. While MULTINEST is designed for evidence evaluation, it also returns the posterior distribution over the model parameters and it is for that purpose that we have used it to date. We have applied the algorithm to the characterisation of several different gravitational wave

data sets — a data set containing two non-spinning supermassive black hole inspiral signals; a data set containing spinning supermassive black hole inspirals and a data set containing bursts from cosmic string cusps. In all three cases, we found that MULTINEST was able to quickly and accurately characterise the main modes of the posterior. Moreover, the algorithm is model-independent in the sense that the specific problem being tackled enters only in the likelihood computation, and does not affect how the live point set is updated.

These promising initial results suggest that it is worth exploring the application of MULTINEST to further problems in gravitational wave data analysis. These include

- Using the evidence computed by the algorithm for model selection, e.g., to identify the number of sources of a particular type in the data stream, or for testing general relativity against alternative theories of gravity.
- Using MULTINEST to search for and characterise extreme-mass-ratio inspiral signals, which are known to have very complex posteriors containing many secondary maxima.
- Applying MULTINEST to the analysis of data from ground-based interferometers (LIGO) — as a search tool, to recover posteriors and for model selection.

The next round of the MLDC [22] will consist of a single data set containing, simultaneously, all of the sources considered separately in round 3. We will apply MULTINEST to the analysis of that data in order to further explore the utility of the algorithm for LISA data analysis.

Acknowledgments

This work used the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (<http://www.hpc.cam.ac.uk/>), provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and we would like to thank Dr. Stuart Rankin for computational assistance. JG's work is supported by the Royal Society. FF is supported by Trinity Hall. The work of SB and AP was supported in parts by DFG grant SFB/TR7 "Gravitational Wave Astronomy" and by DLR (Deutsches Zentrum für Luft- und Raumfahrt). PG is supported by the Gates Cambridge Trust.

References

- [1] Danzmann K *et al.* 1998 *LISA Laser Interferometer Space Antenna, Pre-Phase A Report*, Report MPQ 233
- [2] Arnaud K A *et al.* American Institute of Physics Conference Series, 873, 619 (2006).
- [3] Arnaud K A *et al.*, Class. Quantum Grav. **24** S529 (2007).
- [4] Arnaud K A *et al.*, Class. Quantum Grav. **24** S551 (2007).
- [5] Babak S *et al.*, Class. Quantum Grav. **25** 114037 (2008).
- [6] Babak S *et al.*, Class. Quantum Grav. **25** 184026 (2008).
- [7] Babak S, Gair J R & Porter E K, Class. Quantum Grav. **26** 135004 (2009).
- [8] Skilling J, American Institute of Physics Conference Series, 735, 395 (2004).
- [9] Feroz F, Hobson M P & Bridges M, MNRAS **398** 1601 (2009).
- [10] Feroz F, Marshall P J & Hobson M P, preprint arxiv:0810.0781 (2008).
- [11] Feroz F, Hobson M P, Zwart J T L, Saunders R D E & Grainge K J B, MNRAS **398** 2049 (2009).
- [12] Feroz F, Allanach B C, Hobson M P, Abdus Salam S S, Trotta R & Weber A M, JHEP, 10, 64 (2008).
- [13] Trotta R, Feroz F, Hobson M P, Roszkowski L & Ruiz de Austri R, JHEP, 12, 24 (2008).
- [14] Feroz F, Gair J R, Hobson M P & Porter E K, Class. Quantum Grav. **26** 215003 (2009).
- [15] Veitch J & Vecchio A, Class. Quantum Grav. **25** 184010 (2008).
- [16] Gair J R & Porter E K, Class. Quantum Grav. **26** 225004 (2009).
- [17] Jaranowski P, Królak A & Schutz B F, Phys. Rev. D **58** 063001 (1998).
- [18] Cornish N J & Porter, E K, Class. Quantum Grav. **24** 5729 (2007).
- [19] Cornish N J & Rubbo L J, Phys. Rev. D **67** 022001 (2003); Erratum-ibid. **D67** 029905 (2003).
- [20] Key J S & Cornish N J, Phys. Rev. D **79** 043014 (2009).
- [21] Feroz F, Gair J R, Graff P, Hobson M P & Lasenby A, preprint arxiv:0911.0288 (2009).
- [22] Babak S *et al.*, preprint arxiv:0912.0548 (2009).