Amaldi 9 and NRDA 2011 IOP Publishing
Journal of Physics: Conference Series **363** (2012) 012048 doi:10.1088/1742-6596/363/1/012048

Improving Markov Chain Monte Carlo algorithms in LISA Pathfinder Data Analysis

N Karnesis, M Nofrarias, C F Sopuerta and A Lobo

Institut de Ciències de l'Espai (CSIC-IEEC), Facultat de Ciències, Campus UAB, Torre C5 parells, Bellaterra, 08193 Barcelona, Spain

E-mail: `karnesis@ieec.uab.es`

Abstract. The LISA Pathfinder mission (LPF) aims to test key technologies for the future LISA mission. The LISA Technology Package (LTP) on-board LPF will consist of an exhaustive suite of experiments and its outcome will be crucial for the future detection of gravitational waves. In order to achieve maximum sensitivity, we need to have an understanding of every instrument on-board and parametrize the properties of the underlying noise models. The Data Analysis team has developed algorithms for parameter estimation of the system. A very promising one implemented for LISA Pathfinder data analysis is the Markov Chain Monte Carlo. A series of experiments are going to take place during flight operations and each experiment is going to provide us with essential information for the next in the sequence. Therefore, it is a priority to optimize and improve our tools available for data analysis during the mission. Using a Bayesian framework analysis allows us to apply prior knowledge for each experiment, which means that we can efficiently use our prior estimates for the parameters, making the method more accurate and significantly faster. This, together with other algorithm improvements, will lead us to our main goal, which is no other than creating a robust and reliable tool for parameter estimation during the LPF mission.

1. Introduction

The European Space Agency (ESA) is working on a new space-based Gravitational-Wave (GW) observatory that follows the concept of the NASA-ESA LISA mission [1]. LISA is a constellation of three spacecrafts that will access GW signals at frequencies of around 1 mHz. The scientific working principle of LISA is the detection of picometric relative displacements between pairs of proof masses in nominally geodesic motion, or free fall, induced by passing GWs. Due to the technological complexity of LISA, ESA approved a precursor mission, LISA PathFinder [2, 3], to test the readiness of the main LISA technology. LPF consists in a single spacecraft hosting two proof masses in nominal free fall, whose motions are monitored by means of Mach-Zender laser interferometry. The top level requirement on LPF is usually expressed in terms of relative acceleration between the Test Masses spectral density of noise, and is given by

$$S_{\Delta a}^{1/2}(\omega) = 3 \times 10^{-14} \left[1 + \left(\frac{\omega/2\pi}{3\,\mathrm{mHz}} \right)^2 \right] \mathrm{ms}^{-2}\mathrm{Hz}^{-1/2}, \qquad 1\,\mathrm{mHz} \leq \frac{\omega}{2\pi} \leq 30\,\mathrm{mHz}$$

The LPF mission is expected to be launched in 2014 and its ultimate objective is to demonstrate the purity of the test mass free fall up to the sensitivity expressed above, and to understand the physical origin of the different noise contributions. There are many sources of noise identified (thermal, magnetic, particles of cosmic origin, etc), and properly modeling them requires a careful planning of the measurement sequence, plus the use of suitable analysis tools to process the various data channels. During flight operations the LPF is going to perform a series of experiments with the main goal to fully characterize the physical model of the instrument. Each experiment will determine a set of parameters that are going to be used as additional information for the next experiment in the sequence. Using a Bayesian framework for LPF Data Analysis [7], we are flexible enough to combine knowledge obtained from different experiments. Moreover we are able to use this knowledge as *a priori* information for the next in line parameter estimation experiment, making the algorithm significantly faster. In this work we take advantage of exactly this feature and we present the benefits of using prior information in combination with the MCMC algorithm developed for LPF Data Analysis. First we applied the enhanced algorithm to a simple toy model and then to a more complex LTP model widely used in LTP Data Analysis Operational Exercises [8].

For the LPF Data Analysis, a MATLAB Toolbox has been created [12] by the Data Analysis Team. This package is object-oriented code, based on so called Analysis Objects. The philosophy is to create a highly capable and flexible tool, with strong history tracking capacity to keep memory of the work done on any data segment, or derived object by anybody using the tool. The present work has been implemented as a part of the LTP Data Analysis toolbox (LTPDA).

2. A Bayesian Framework

Markov Chain Monte Carlo (MCMC) is a subset of Monte Carlo methods that involve a Markov process in which a sequence of samples $\{\theta_n\}$ is generated, each sample θ_n having a probability distribution that depends on the previous one, θ_{n-1} [5, 6]. This means that new samples in the parameter space are generated using a Markov Chain mechanism and in each jump (or step) in the parameter space, the likelihood is calculated. After a sufficient number of samples, one can investigate the shape of the likelihood surface. More precisely, one can be based to some prior information about the system and update this information based on the observed data $\vec{\mathbf{x}}$. This procedure is based on the well-known Bayes Theorem:

$$\pi(\vec{\theta}|\vec{\mathbf{x}}) = \frac{f(\vec{\mathbf{x}}|\vec{\theta})p(\vec{\theta})}{f(\vec{\mathbf{x}})} \tag{1}$$

where $\vec{\theta}$ is a parameter vector, $f(\vec{\mathbf{x}}|\vec{\theta})$ is the *Likelihood* of the dataset $\vec{\mathbf{x}}$, given the parameters $\vec{\theta}$ and $\pi(\vec{\theta}|\vec{\mathbf{x}})$ and $p(\vec{\theta})$ are the *posterior* and the *prior* distributions respectively. Note that the *evidence* $f(\vec{\mathbf{x}})$ is independent of the parameters $\vec{\theta}$ and it is often neglected in parameter estimation algorithms, as it serves only as a normalising constant:

$$\pi(\vec{\theta}|\vec{\mathbf{x}}) \propto f(\vec{\mathbf{x}}|\vec{\theta})p(\vec{\theta}) \tag{2}$$

As already mentioned we need to successfully integrate the posterior distribution. A most used MCMC algorithm is the *Metropolis algorithm.* The Metropolis algorithm can be summarized in the following steps. First, an initial set of parameters $\vec{\theta}_0$ is chosen. Then a new point in the parameter space is drawn from a multivariate Gaussian distribution. The covariance matrix Σ of the multivariate Gaussian is extracted from the calculation of the inverse of the Fisher Information Matrix [7]. The FIM is calculated once in a point on the parameter space that is our first "guess" of the parameters and the proposal distribution is constant during the MCMC run. In the newly created point, the likelihood is calculated and then the *acceptance probability* α is evaluated. The acceptance probability,

$$\alpha = \min\left(1, \frac{f(\vec{\mathbf{x}}|\vec{\theta})p(\vec{\theta})q(\vec{\theta}_0|\vec{\theta})}{f(\vec{\mathbf{x}}|\vec{\theta}_0)p(\vec{\theta}_0)q(\vec{\theta}|\vec{\theta}_0)}\right) \tag{3}$$

will determine whether if the new point is accepted or rejected by the algorithm. If the point is accepted, we set $\vec{\theta}_0 = \vec{\theta}$ and propose another point from the parameter space. The proposal densities q are vanished from equation 3 since they are symmetrical. This procedure is repeated until we converge to a set of parameters $\vec{\theta}_{\mathrm{MAP}}$ that maximize the posterior distribution.

2.1. MCMC and the LTPDA

From Eq. (2) one can see that prior Probability Density Functions (PDFs) can play an important role to the evolution of the algorithm. Until now, in the MCMC developed for the LTPDA toolbox [7], the prior information about the parameters was limited to flat PDFs. This translates to a "closed box" in the parameter space, where we assume that the $\vec{\theta}_{\mathrm{MAP}}$ is located. The nature of LPF mission allows us to use any previous estimates for the parameters for the next experiment in the sequence. This makes the introduction of priors into the algorithm an essential improvement. The result of the first experiment is a set of PDFs for the parameters. These posterior PDFs can be used as prior information for the next experiment and what we expect theoretically is faster convergence to $\vec{\theta}_{\mathrm{MAP}}$. The MCMC LTPDA algorithm was improved taking into account the prior PDFs, as in Eq. (3). The next step was to study what exactly is our gain in terms of time and accuracy and apply our findings to a more complete and complex LTP model.

3. Numerical Investigations

In order to discover and verify the benefits of using results from previous experiments as prior knowledge for the following ones, we designed a series of studies using a toy model. A simple one-dimensional low pass filter was chosen, since the results from MCMC runs were obtained faster. In section 3.2 we applied the same to a LTP model.

3.1. Special case: Toy model

A simple toy model was used to investigate the behavior of the algorithm. The model is a first-order low-pass filter with cutoff frequency $f_0 = 0.01\,\mathrm{Hz}$. A data set and noise were created after filtering a single frequency sinusoidal input. Then we used the simple model and inputs to estimate the supposedly unknown cutoff frequency f_0 of the low pass filter toy model. The posterior PDF of f_0 estimated is then imported to the next experiment, of the same type as the first one, as prior PDF. The idea behind this is to illustrate any effect to the speed of convergence of the algorithm. In order to properly estimate the time of convergence, we estimate the unknown f_0 fifty times for each configuration. For the first runs we used flat prior PDFs and we were able to fit a Gaussian curve to the posterior PDFs computed and use it as a prior PDF for the following runs. Following, we estimated f_0 using the Gaussian priors with their variance σ multiplied by a constant in order to see the exact effects of the prior on the convergence of the algorithm (the σ is in the order of 10^{-4}). In particular we did:

- normal runs with flat priors.
- Runs with priors with original σ.
- Runs with priors $50 \times \sigma$
- Runs with priors $500 \times \sigma$
- Runs with priors $1000 \times \sigma$

In figure 1, the complete progress of the chains is presented. In these chains the rejected points are also included. During the first burn-in phase, new points are proposed with the aim to explore aggressively the parameter space. If a point is far away from the area defined from the σ of our prior probability, then the value of the logarithm of the acceptance ratio is dominated

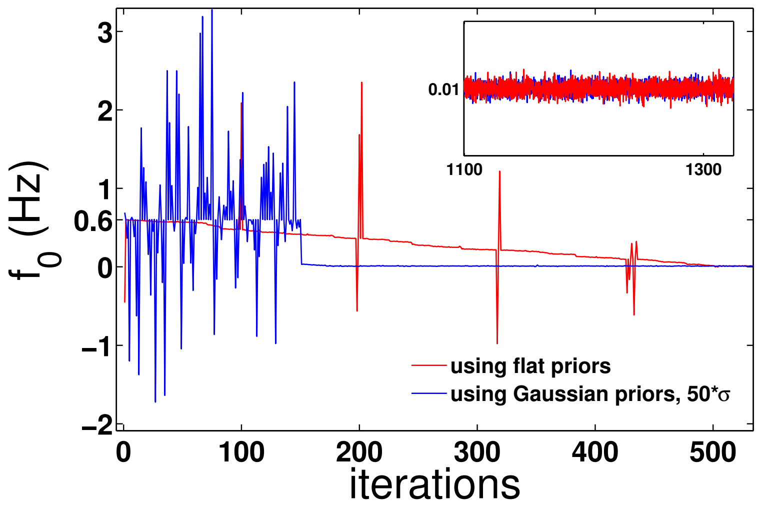

Figure 1. Graph of the chains *including the rejected points* from one of the fifty runs for two σ configurations. In the embedded figure the chains after convergence are displayed.

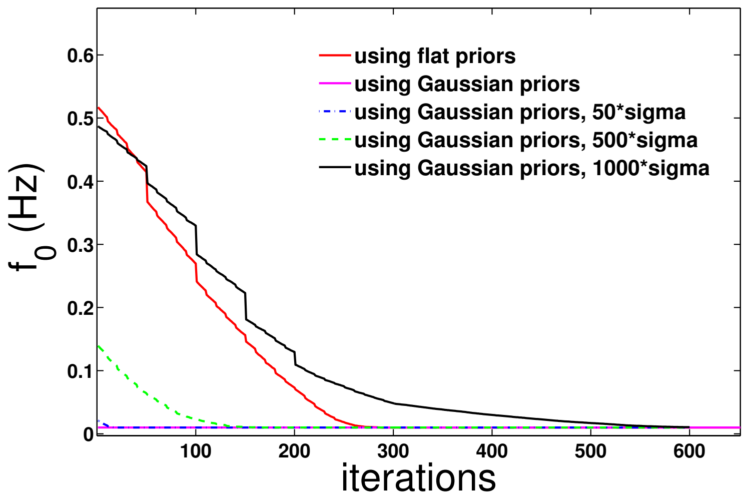

Figure 2. For fifty runs for each σ we can see the mean path of the chains of the accepted points.

by the very large terms of $log(p(\vec{\theta}))$ and the log-likelihood values are negligible compared to them. This results to a rejection of the new point proposed and while the acceptance ratio is computed, the costly calculation of multiplication of large matrices can be skipped. Now, this first searching phase becomes much more rapid. In figure 2, we can clearly see that as the σ gets narrower, the path of the chains becomes shorter in their way to the value of f_0 that maximizes the posterior.

3.2. Application to a LTP model

In order to understand the future LPF mission, a number of models of the LTP [10] have been implemented by the LTP DA team. One widely used for operational exercises [8, 9] is depicted in figure 3. It represents the experiment as a close-loop system, where noise contributions and cross-couplings are expresed as a set of transfer functions in frequency domain. This particular

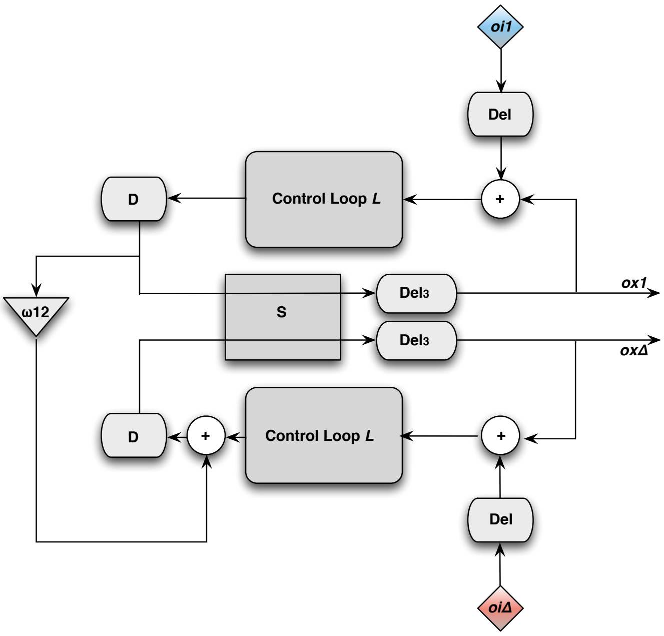

Figure 3. Illustration describing the simplified LTP model. This model describes the LTP operating in the Main Science Mode [11]. See text for a description of terms appearing in the picture.

model is a simplified model for the LTP, thus, only two heterodyne interferometers are taken into account: one measuring the distance between the two test masses ($o_{x\Delta}$) and one measuring the distance between test mass one (TM1) and the spacecraft (represented here as o_{x1}). The complete system, after the assumption that all noise terms are independent from the unknown parameters, can be expressed by:

$$\vec{o} = \mathbf{H}_s(\vec{\theta})\vec{o}_i + \vec{n} \tag{4}$$

This is justified as we are working in the high signal-to-noise ratio regime and consequently the dependance of the noise from the parameters is negligible. The terms appearing in Eq. (4) are: $\vec{o}$ is the vector of the interferometer readouts, $\vec{o}_i$ is vector of the injected signals, $\vec{n}$ is the overall instrument noise and $\mathbf{H}_s(\vec{\theta})$ is the complete transfer function describing the LTP. Each of the elements of Eq. (4) read as:

$$\vec{o} = \begin{pmatrix} o_{x1} \\ o_{x\Delta} \end{pmatrix}, \qquad \vec{o}_i = \begin{pmatrix} o_{i1} \\ o_{i\Delta} \end{pmatrix}, \qquad \vec{n} = \begin{pmatrix} n_{x1} \\ n_{x\Delta} \end{pmatrix}$$

$$\mathbf{H}_s(\vec{\theta}) = (\mathbf{D} \times (\mathbf{Del_3} \times \mathbf{S}) + \mathbf{L})^{-1} \times \mathbf{L} \times \mathbf{Del}, \tag{5}$$

where $\mathbf{H}_s(\vec{\theta})$ is the 4×4 matrix transfer function of the complete system. $\mathbf{H}_s(\vec{\theta})$ is composed of $\mathbf{D}$; the matrix describing the dynamics of the system; the $\mathbf{S}$ matrix, which describes the interferometric read-out and possible construction imperfections; the $\mathbf{Del_3}$ and **Del**, describing time delays; and finally the **L**, which describes the *Control Loop* of the LTP. The Control Loop contains three main subsystems: the inertia decoupling matrix, a matrix with the Drag-Free Attitude and Control System transfer functions and the electrostatic actuator gain and delays. A more detailed description can be found in [9].

The steps followed here are the same as in the Operational Exercise 6 [9]. A frequency sweep of sinusoidal signals was injected to the system forming two independent investigations. For the first investigation we injected a signal o_{i1} into the drag-free loop and for the second we injected a signal $o_{i\Delta}$ into the low-frequency suspension loop. The Operational Exercise 6 was designed with the aim of fully characterizing all the possible dynamical responses of the system-model of the LTP using parameter estimation methods developed by the LTPDA team. These particular investigations and model were chosen in order to perform the same complex analysis for LTP as in operational exercises and prove that the use of prior information applies to up-to-date parameter estimation methods.

As in the toy model, we used previous estimates of the parameters as prior information for our next experiments. The PDFs of the parameters of the first estimation were fitted with Gaussian curves and multiplied by a factor before we input them into the algorithm. The parameters to estimate are: ω_1, the stiffness coupling of TM1 to S/C, ω_{12}, the difference between stiffness coupling TM1 to S/C and TM2 to S/C, A1, the thrusters actuation gain, A2, the electrostatic actuation gain, DT1, the delay in the application of the thrusters actuation, DT2, the delay in the application of the electrostatic actuation and δ_{21} - the cross coupling between the interferometer channels.

In Figure 4 we can clearly see the difference of evolution of the MCMC chains for four out of seven total parameters of the LTP model using flat and Gaussian prior PDFs. While flat prior PDFs give little information about the parameter space that maximize the posterior, the Gaussian prior PDFs rapidly force the chains to explore the small area of interest. But in order to study the benefits of this practice we measured the time the algorithm needed until it achieved convergence to $\vec{\theta}_{\mathrm{MAP}}$. The algorithm was executed five times for each configuration and the mean time of execution was calculated. Basically, following this practice we are able to reduce the first burn-in period in terms of samples. Depending on the problem, the burn-in period could range between 3000 to 5000 samples. In this particular investigation it is reduced by 2000 samples, thus using a total of 5000 and 7000 for each case respectively.

Table 1. Mean computation time for both methods.

Method	mean time (s)
using flat priors	4602
using normal priors	2528

4. Results and future work

In table 1 the results of this investigation are summarized. The numbers presented are strongly correlated to the machine operating the toolbox, but it is clear that we gain almost a factor of two in terms of time. For the MCMC algorithm, faster convergence means that we can have efficient estimation of the parameters with fewer sample points. This is exactly what happens when we use Gaussian prior PDFs. Without *a priori* information we have to explore a relatively large

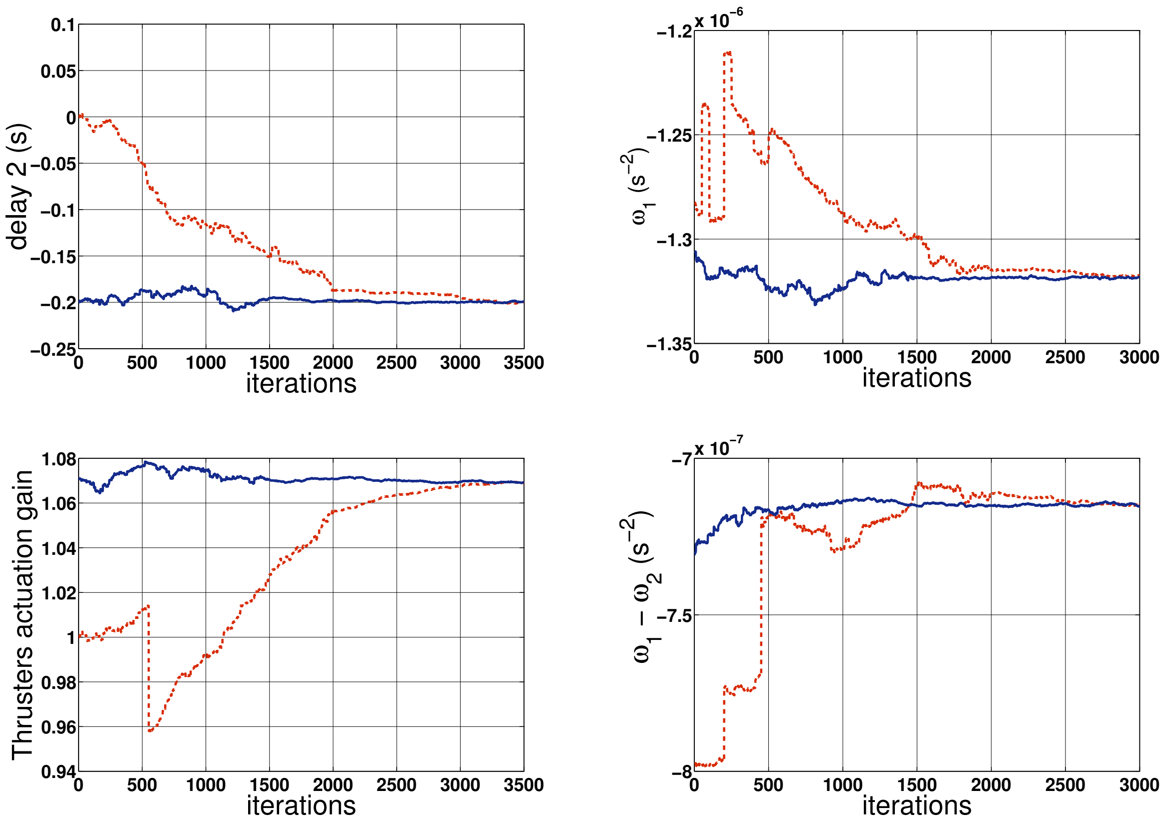

Figure 4. MCMC chains evolution for the estimation of the thrusters actuation gain A1, the delay2, the ω_1 and the $\omega_1 - \omega_2$ of the LTP model. The chains of the rest of the parameters of the model follow the same pattern. Here only the burn-in and cooling parts of the MCMC chains are presented. The dashed lines represent the chains when we use flat PDFs as prior knowledge, while the solid lines represent the chains when we use Gaussian prior information.

area of the parameter space. This means that again, a large number of samples has to be spent in the burn-in and cooling periods of the MCMC chains. These samples are then discarded and the remaining parts of the chains are used for the estimation of the parameters. The convergence of the chains is empirical, since, it is considered that given sufficient time, the MCMC will almost always converge to the $\vec{\theta}_{MAP}$. Also in the majority of system identification experiments of LPF, the Signal to Noise Ratio is huge and the MCMC reaches the MAP area quickly. In our case, the burn-in and cooling periods required more than half of the total samples gathered by the algorithm and usually the final 2000 to 3000 samples are used to make an estimation of the parameters. In our study the burn-in and cooling periods become significantly smaller, since the area to explore in the parameter space reduces to a small fraction of the original one. Although this practice seems to be consistent with the expected results we did not take in to account the correlation between the parameters in the calculation of the prior PDFs. That means that we used only the information contained in the diagonal of the covariance matrix extracted from the first estimation of the parameters. A next step is to include all the elements of the covariance matrix for a more complete calculation of the prior probability. Furthermore, a more complete version of the algorithm will take into account the fact that the narrow Gaussian priors affect the computation of the Fisher Information matrix and as a result the covariance matrix used in the proposal distribution. After completing this first step of enhancing the LTPDA

Journal of Physics: Conference Series **363** (2012) 012048 doi:10.1088/1742-6596/363/1/012048

MCMC algorithm we plan further improvements and adaptations, like tuning the proposal distribution [13] and implementing various types of annealing and parallelizing.

We have shown how prior information can be used for parameter estimation and how it affects the time consumed by the algorithm. A simple model was implemented and a series of experiments were carried out. The results showed that *a priori* information provide accurate estimates using fewer samples. As expected, the same results were obtained for a more complex model describing the LTP onboard LPF. The time consumed by a statistical tool like MCMC is significantly reduced, making the method more suitable for Data Analysis between the experiments during the mission. This is a first step towards a more complete parameter estimation tool for LPF.

Acknowledgments
The authors acknowledge support from contract AYA2010-15709 of the Spanish Ministry of Science and Innovation. NK acknowledges support from the FI PhD program of the AGAUR (Generalitat de Catalunya). MN acknowledges support from a JAEDoc contract of the National Spanish Research Council (CSIC). CFS acknowledges support from the Ramón y Cajal Programme of the Ministry of Education and Science of Spain and by a Marie Curie International Reintegration Grant (MIRG-CT-2007-205005/PHY) within the 7th European Community Framework Programme.

5. References

[1] Danzmann K, and LISA Science Team, 2003 *Adv. in Space Res.*, **32**, Issue 7, 1233-1242
[2] Armano M *et al* 2009, *Class. Quantum Grav.* **26**, 09400.
[3] Dolesi R *et al* 2003, *Class. Quantum Grav.* **20** S99
[4] Kendall W S, Liang F, Wang J S, *Markov Chain Monte Carlo Innovations and Applications*, (Singapore:World Scientific Publishing)
[5] MacKay D J C, 2003, *Information Theory, Inference, and Learning Algorithms*, (Cambridge: C.U.P.)
[6] Andrieu C, Doucet A and Jordan M I, 2003, *Mach. Learn.*, **50**, 543
[7] Nofrarias M, Röver C, M Hewitson, A Monsky, G Heinzel and K Danzmann 2010 *Phys. Rev.* D **82** 122002
[8] F Antonucci *et al*, 2010, *Class. Quantum Grav.* **28**, 094006
[9] Nofrarias M, Ferraioli L, Congedo G, 2011, *Tech. Rep.* S2-AEI-TN-3070, Issue 1
[10] Hewitson M, Grynagier A, Diaz-Aguilo M, *Tech. Rep.* S2-AEI-TN-3069, Issue 1
[11] Fichter W, Gath P, Vitale S and Bortoluzzi D, 2005 *Class. Quantum Grav.* **22** S139
[12] Hewitson M *et al* 2009, *Class. Quantum Grav.* **26** 094003
[13] Röver C 2010 *J. Phys.: Conf. Series* **228** 012008