Recovering Cosmic String Cusp Waveforms in Mock LISA Data Challenge 3.4

Joey Shapiro Key and Neil J. Cornish
Department of Physics, Montana State University, Bozeman, MT 59717
(Dated: April 27, 2009)

We employ a parallel tempered Markov Chain Monte Carlo (MCMC) approach to detect and characterize the signals from cosmic string cusps in simulated data from the third round of Mock LISA Data Challenges (MLDCs).

I. INTRODUCTION

Round three of the Mock LISA Data Challenge includes training and blind data sets with simulated gravitational wave signals from cosmic string cusps [1]. We present here our search technique, our results from the training data, and our results from the blind data set.

II. MCMC SEARCH TECHNIQUES

Our search is based on a Markov Chain Monte Carlo method [2, 3]. We employed several techniques to ensure rapid exploration of the full parameter space: local coordinate transformations to uncouple the parameters; moves that exploit symmetries of the likelihood surface to encourage jumps between local maxima; and parallel tempering to encourage wide exploration of the posterior [4]. Details of our search technique can be found in Ref. [5].

The cusp gravitational wave signal [6–9] was parameterized by $\vec{x} \rightarrow \{\ln A, t_*, \ln f_{\rm max}, \theta, \phi, \psi\}$, and the priors were taken to be uniform in these quantities, save for θ, where a uniform sky distributions is given by $\Pi(\theta) = \frac{1}{2}\sin(\theta)$.

A. Detector Symmetry Based Proposals

Any burst signal with duration less than a day produces a response in the LISA detector that can be well approximated by a stationary antenna. Such signals will suffer from a degeneracy such that sky locations related by a reflection in the plane of the detector with produce an *identical* response. There are additional symmetries in the low frequency limit that result from 120° rotations in the plane of the detector. These symmetries are broken at higher frequencies by the slightly different arrival times of the gravitational waves across the LISA array. Rotations thus produce two sets of secondary maxima. We employ proposal distributions that propose jumps to these symmetry locations as part of our MCMC search.

B. Parallel Tempering

Global exploration of the parameter space is enhanced by creating a set of parallel chains with likelihood surfaces at different "Temperatures" T such that

$$\mathcal{L}_i(\vec{x}) = \mathcal{L}(\vec{x})^{1/T_i} \quad . \tag{1}$$

Chains that explore surfaces with $T \gg 1$ tend to take bigger steps since the contrast between maxima and minima is decreased, and this encourages wider exploration of the parameter space [10, 11]. The chains can exchange parameters according to the Hastings ratio

$$H_{PT} = \frac{\mathcal{L}_a(\vec{x}_b)\mathcal{L}_b(\vec{x}_a)}{\mathcal{L}_a(\vec{x}_a)\mathcal{L}_b(\vec{x}_b)} \quad , \tag{2}$$

for chains with temperature T_a and T_b and parameters $\vec{x}_a$ and $\vec{x}_b$, respectively.

We implement the parallel tempering method for N_C chains with the T values given by

$$T_i = (\Delta T)^{i-1} \tag{3}$$

where

$$\Delta T = (T_{\rm max})^{\frac{1}{N_C - 1}} \quad . \tag{4}$$

Only the $T = 1$ chain samples the true PDF and is used to produce the parameter histograms. We typically used 20 chains and a maximum heat of $T \sim 100$.

III. THE MOCK LISA DATA CHALLENGE

Challenge 3.4 is comprised of a month long data set (2^{21} samples with 1 second sampling) with cosmic string cusp waveforms injected with a Poisson event rate of five events per month. This is the first MLDC data set with non-symmetric instrument noise. The cusp burst sources can be found using a symmetric approximation for the noise (leading to a small systematic bias in the recovered parameters), or the source parameters and the individual noise levels can be fitted simultaneously in the search.

The time of arrival at the solar system barycenter (t_*) is highly correlated with the sky location of the source. A search for t_* leads to poor determination of the time of arrival of the burst due to the inherently poor resolution of the sky location. A better choice of variable is the time of arrival at the guiding center of the LISA constellation ($t_\triangle$). The detector time of arrival is not as correlated with the sky location parameters, resulting in better conditioned Fisher Information Matrices to drive the local jumps of the Markov Chain.

Source	SNR1	t_* (sec)	SNR2	t_* (sec)
3.4.0	53.54	386501.8	53.99	386502.6
3.4.1	21.46	1889293	21.27	1889293
3.4.2	31.07	1864208		
3.4.3	73.87	2059895	76.58	2059896
3.4.4	14.07	1498159	14.12	1498164

TABLE I: The triggers produced by a search of the MLDC 3.4 training data. All but one of the sources was detected in both passes through the data (one signal happened to straddle a data segment boundary). The time of the trigger is listed, along with the recovered value for SNR.

A. Training Data Results

The triggers found by our search in the training data are shown in Table I and the MAP parameters for each source are shown in Table II.

The training data was analyzed without reference to the answer key so as to mimic the steps that will be taken to analyze the blind challenge data. The parallel tempering technique takes care of both detection and characterization, so the analysis does not have to be be broken up into distinct stages. On the other hand, running the search on the full $\sim 2\times 10^6$ seconds of data to find signals with duration $\sim 10^3$ seconds is not very efficient, so we adopted the strategy of dividing the full data set into 64 segments of length 32,768 seconds. Time domain filters were used to limit spectral leakage, and the finite response of these filters meant that signals in the first and last $\sim 10\%$ of each segment had to be discarded. To ensure full coverage, a second pass was performed using segments offset from the first by $16,384$ seconds.

The first stage of the analysis was to search each data segment using $N_C = 12$ chains with $\Delta T = 1.55$ and $N = 10,000$ iterations. A simple SNR threshold was used to decide if a source had been found in the data segment. Triggers with $\text{SNR} = (s|h)^{1/2} > 8$ were recorded for further analysis (the loudest noise triggers had SNR < 6). If a trigger was found the signal was regressed from the data and the search repeated (in other words the search is sequential rather than simultaneous).

The initial search did not fit for instrument noise levels, but found all five signals in the training data and recovered the source parameters to good accuracy (Table I). Since the segmented data is searched twice, we expect to find each source twice, but one trigger happened to fall near the boundary between segments and was thus discarded. The source was found on the offset pass.

We evaluated our results by calculating the ρ statistic, and the correlation between the recovered template and the data (the latter could only be checked for the training data as the answer key for the blind data is not yet available).

In terms of the usual noise weighted inner product:

$$(a|b) = \frac{2}{T_{obs}} \sum_{A,E,T} \sum_{f} \frac{a^*_\beta(f) b_\beta(f) + a_\beta(f) b^*_\beta(f)}{S^\beta_n(f)} \quad (5)$$

where T_{obs} is the observation time and $S_n(f)$ is the one-sided noise spectral density in each channel, the rho statistic is defined:

$$\rho = \frac{(s|h)}{\sqrt{(h|h)}}, \quad (6)$$

where s is the data and h is entry template. The correlation is defined as

$$\text{Correlation} = \frac{(h|s)}{\sqrt{(h|h)(s|s)}}. \quad (7)$$

We only use data from withing a window of 1024 seconds around the burst when computing correlations. When noise is present the correlation is not a very useful measure of performance. The "theoretical" SNR of the entry template is

$$\text{SNR} = (h|h)^{1/2}\,. \quad (8)$$

A good solution will have $\rho \sim$ SNR.

Source	θ (rad)	ϕ (rad)	ψ (rad)	A	t_* (sec)	$f_{\max}$
3.4.0	-0.2485	5.589	0.3997	1.986e-21	386843.98	2.37e-3
MAP	3.301e-2	5.277	0.3464	3.647e-21	386729.38	2.27e-3
3.4.1	-1.181	4.790	5.205	1.073e-21	1889234.54	1.157
MAP	4.232e-2	4.797	1.959	1.440e-21	1889147.75	4.06e-2
3.4.2	-0.9337	0.6772	3.981	6.604e-22	1864491.28	0.4642
MAP	0.4327	3.635	2.468	5.998e-22	1863759.94	2.57e-2
3.4.3	0.2391	1.090	3.541	2.647e-21	2060273.07	1.15e-2
MAP	-0.1731	2.240	2.968	1.079e-20	2059769.79	1.05e-2
3.4.4	-1.030	1.156	1.099	3.420e-22	1498329.95	2.277
MAP	-0.6732	2.703	2.794	5.815e-22	1497869.97	2.54e-2

TABLE II: The MLDC training data sources with the injected parameters and the recovered MAP parameters found with our search.

The results in Table III show that the templates constructed with the MAP parameters from our search have a high correlation with the noise-free training data. The lisatools software was used to produce the templates used to make this comparison. As expected, the value of the ρ statistic is very close to the theoretical SNR, and we use this as an indication of how well our technique has determined the source parameters in the blind search, when there is no noise-free data available.

B. Blind Data Results

The same methods employed in the search for cosmic string signals in the training MLDC data were used to

MLDC 3.4.0	SNR	ρ	Correlation
Noise-Free	41.58	40.25	0.999883
With Noise	41.58	40.03	0.872481

MLDC 3.4.1	SNR	ρ	Correlation
Noise-Free	21.08	20.93	0.998167
With Noise	21.08	23.37	0.627839

MLDC 3.4.2	SNR	ρ	Correlation
Noise-Free	29.26	29.15	0.999483
With Noise	29.26	32.44	0.763786

MLDC 3.4.3	SNR	ρ	Correlation
Noise-Free	74.74	74.91	0.999897
With Noise	74.74	73.45	0.948604

MLDC 3.4.4	SNR	ρ	Correlation
Noise-Free	13.08	12.78	0.996606
With Noise	13.08	11.79	0.648446

TABLE III: The MLDC training data best fit parameters recovered from the search produce waveforms that can be compared with the noise-free MLDC data sets and the full MLDC data sets to determine how well each source was recovered.

find the signals in the blind data. Three sources were found in the month long data set. The triggers are listed in Table IV and the recovered MAP parameters are listed in Table V. The evaluation of the recovered template can be found in Table VI with comparable ρ and SNR values indicating a good match to the data.

	SNR1	t_* (sec)	SNR2	t_* (sec)
Source 0	40.461692	599202.3	40.764385	599242.1
Source 1	33.094462	1072929	33.071948	1072929
Source 2	43.264396	1603018		

TABLE IV: The triggers produced by a search of the MLDC 3.4 blind data with the time of each trigger and the recovered SNR. One source was found on only one pass through the data since it straddled the segment boundary.

	θ (rad)	ϕ (rad)	ψ (rad)	A	t_* (sec)	$f_{\max}$
Source 0	0.3094	3.926	4.552	9.912e-22	599287.64	Nyquist
Source 1	-0.3233	3.934	4.957	2.763e-21	1072739.28	1.056e-3
Source 2	0.2325	5.899	5.919	1.512e-21	1602943.85	Nyquist

TABLE V: The recovered MAP parameters for the MLDC blind search.

MLDC source 0	SNR	ρ	Correlation
With Noise	40.34	42.09	0.748116

MLDC source 1	SNR	ρ	Correlation
With Noise	20.16	19.62	0.815867

MLDC source 2	SNR	ρ	Correlation
With Noise	42.71	42.80	0.679256

TABLE VI: The MLDC blind data best fit parameters recovered from the search produce waveforms that can be compared with the data to test how well each source was recovered.

IV. CONCLUSION

A matched filter analysis using parallel tempered Markov Chain Monte Carlo techniques can both detect and characterize the gravitational wave signals from cosmic string cusps in simulated LISA data as demonstrated by the Mock LISA Data Challenge results.

V. ACKNOWLEDGEMENTS

This work was supported by NASA grant NNX07AJ61G.

[1] *The Mock LISA Data Challenges: from Challenge 1B to Challenge 3*, 12th (Gravitational Wave Data Analysis Workshop, Cambridge, MA, 2008).

[2] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, Journal of Chemical Physics **21**, 1087 (1953).

[3] W. Hastings, Biometrika **97**, 57 (1970).

[4] N. J. Cornish and J. Crowder, Phys. Rev. D (2005).

[5] J. S. Key and N. J. Cornish, Phys. Rev. **D79**, 043014 (2009).

[6] X. Siemens and K. Olum, Phys. Rev. D **085017**, 68 (2003).

[7] X. Siemens, J. Creighton, I. Maor, S. R. Majumder, K. Cannon, and J. Read, Phys. Rev. D **105001**, 73

(2006).
[8] M. R. DePies and C. J. Hogan, Phys. Rev. D **75** (2007).
[9] X. Siemens, J. Key, and N. Cornish, in preparation (2008).
[10] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: A Comprehensive Approach with Mathmatica Support* (Cambridge University Press, 2005).
[11] P. Gregory, Astrophys. J. **631**, 1198 (2005).