Technical Note on EtfAG Entry for MLDC Round 1B.3

Jonathan R Gair*

Institute of Astronomy, Madingley Road, CB3 0HA, Cambridge, UK and Max-Planck-Institut für Gavitationsphysik, Am Mühlenberg 1, 14476, Potsdam, Germany

Ilya Mandel†

Department of Physics and Astronomy, Northwestern University, 2131 Tech Dr., Evanston, IL 60208

Linqing Wen‡

Max-Planck-Institut für Gavitationsphysik, Am Mühlenberg 1, 14476, Potsdam, Germany and School of Physics, M013, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

(Dated: December 3, 2007)

We carried out an analysis of Mock LISA Data Challenges 1B.3.1-1B.3.5 using time-frequency techniques. We were able to detect EMRI signals in all of these data sets, and estimated 9 parameters for each challenge data set: the two masses, eccentricity and frequency at $t = 0$, SMBH spin, spin-orbital angle λ, two sky direction angles, and the angle κ between the SMBH spin vector and the line of sight to the source. We also independently estimated the plunge time, but this is determined by the other intrinsic parameters and was used to improve the determination of those parameters. In this note, we focus on the changes we have made since the last round of the MLDC, 1.3.*. We describe the changes in our track detection algorithm, then the improvements in our parameter estimation routines, and finally comment on some details of individual challenges.

I. TRACK DETECTION

A. Chirp Track Search

In the previous round of the MLDC, we used the Hierarchical Algorithm for Clusters and Ridges (HACR) to detect tracks from EMRI signals in the time-frequency spectrogram [1]. HACR [2] is a good tool for detecting clusters of bright pixels in a spectrogram. However, HACR does not use all of the available information about the signal. Specifically, we expect the tracks in the time-frequency spectrogram to be chirping curves characterized by three parameters: frequency f and its first two time derivatives $\dot{f}$ and $\ddot{f}$. (The third time derivatives of frequency along the tracks are vanishingly small for low-eccentricity signals like those in challenges 1B.3.1-1B.3.3, and can be neglected for challenges 1B.3.4 and 1B.3.5 if sufficiently short sections of the tracks are fitted.) Moreover, the first and second time derivative of frequency must be positive. This additional information can help to detect tracks. In fact, when we scanned the time-frequency spectrogram by eye, we found some tracks that HACR didn't report, because we intuitively applied this additional information about the expected track shape when visually searching for tracks.

Therefore, for this round of MLDC, we built a Chirp-based Algorithm for Track Search (CATS) to aid in track detection. This algorithm works as follows:

- Construct A and E channels from the challenge data stream.
- For each channel, construct a spectrogram by dividing the data into time segments of equal duration, then Fourier transforming the data in each time segment after multiplying it by a Hanning window to reduce edge effects. The time segment durations were 32×10^5 seconds for challenge 1B.3.1, 8×10^5 seconds for challenges 1B.3.2 and 1B 3.3, and 2×10^5 seconds for challenges 1B.3.4 and 1B 3.5.
- Normalize the two spectrograms by dividing the signal power in each pixel by the expected LISA noise power spectral density, then construct a joint spectrogram by adding the two normalized spectrograms.

*Electronic address: `jgair@ast.cam.ac.uk`
†Electronic address: `ilyamandel@chgk.info`
‡Electronic address: `lwen@aei.mpg.de`

- Define the starting and ending time of a chirp search. Construct a three-dimensional grid of parameters in the $(f, \dot{f}, \ddot{f})$ space. For each point in the parameter space, build a potential track. Then add up the power in all pixels along this track between the starting and ending time of the search to obtain the total track power. (This is, effectively, a variation on a template-based matched-filtering search in the $(f, \dot{f}, \ddot{f})$ space.)
- Find the brightest track and claim it as a detection. Set the power in all pixels along this track in the spectrogram to large negative values in order to keep future detected tracks from intersecting the current track. Repeat this step until all tracks are detected.

For this round, we did not attempt to search the whole spectrogram using this technique, nor did we search over the entire parameter space of $(f, \dot{f}, \ddot{f})$. Instead we used the HACR algorithm again to identify clusters in the spectrogram arising from the presence of signals in the data stream, and the approximate shapes of these tracks. We then targeted those areas for follow-up with the CATS algorithm. At present, we have also not set robust thresholds for track detections. In the future, the thresholds will be set based on a desired value of the false alarm probability. For this round, we set thresholds by hand, but performed a number of "sanity checks" to decide whether a given track detection was to be believed. These checks included:

- Compare the detected tracks with a visual observation of the spectrogram.
- Check if the tracks have sidetracks due to FFT bleeding, i.e., tracks that are parallel to them but one pixel off on each side.
- Check if there is a clear harmonic structure typical to EMRIs, with sidebands caused by orbital-plane precession: several tracks that are nearly parallel (same $\dot{f}$) with a fixed offset from each other.
- Check if the track parameters are reasonable: e.g., if presumed sections of a track are found at early and late times, the frequency and frequency derivative at the later time should be greater than at the earlier time.

In challenges 1B.3.*, CATS successfully found all tracks that were detected by HACR or could be made out by eye. For some of the challenges, CATS also found additional sidebands

We estimated the systematic uncertainties of CATS by comparing the results of track searches using different time steps and different gridding of the parameter space. These comparisons suggest uncertainties on the order of $\Delta f \sim 10^{-7}$ Hz, $\Delta \dot{f} \sim 10^{-14}$ Hz/s, and $\Delta \ddot{f} \sim 10^{-21}$ Hz/s^2. Of course, the inherent errors in chirp parameter estimation of low-SNR signals, particularly for short tracks, could be much greater than these systematic uncertainties.

B. Radon Transform Search

We have also been developing and testing another time-frequency technique that utilizes the modified Radon transform. For this search, we sum power in pixels along expected tracks in the time-frequency spectrogram of a given source. The expected tracks are calculated by integrating the known evolution equations for EMRIs for a given set of initial parameters. Tracks of the harmonics and sidebands of orbital frequencies are included in the summation, and we search through all possible parameter space for the maximum signal-to-noise ratio. This technique was not fully developed at the time of submission, so although the results were compared to those results obtained from the CATS algorithm, only the latter results were reported for this Challenge. In the future, further improvement of the Radon transform algorithm will be to needed in order to overcome problems of getting stuck in local maxima of the parameter space. We are also investigating similar techniques using the Hough transform which sums over number counts (instead of powers) of bright pixels along the track. A search using the Radon or Hough transform normally take minutes to complete on a laptop when a good initial guess has been made, or hours starting from a poor guess.

II. PARAMETER EXTRACTION

Our parameter extraction from the CATS tracks consisted of two steps. In the first step, which was described in more detail in the technical note accompanying the previous submission and in [1, 4], we made rough estimates of the intrinsic parameters based on the track shape only and of the sky position using the power variation along the track [1, 3]. If at any time during the observation two harmonics of the orbital frequency are detected, plus a sideband of one of those harmonics arising due to the orbital plane precession, then we can measure the three fundamental frequencies at that time. The orbital frequency, ν, is an intrinsic parameter and, for a choice of eccentricity, e, and orbital inclination, λ, the orbital plane precession frequency, f_α, determines the black hole spin, S, while the perihelion precession frequency, f_γ, determines the central black hole mass, M. If the derivative of a harmonic is also measured, then the compact object mass, m, can also be determined for specified e and λ. If we have an estimate of the plunge time, t_p, and an estimate of at least one of the frequency components at another time, it is possible to iterate over e and λ to obtain a crude estimate of the intrinsic parameters.

For this challenge, we also made use of the relative power in the sidebands of a given harmonic of the orbital frequency. These sidebands arise from the orbital plane precession. From physical arguments, the relative power in these sidebands can only depend on the intrinsic parameters and on the angle, κ, between the line of sight to the source and the spin axis of the black hole. It is possible to write the Barack and Cutler waveform in the solar system barycentre as an expansion in harmonics, $f_i = n_i \nu + 2 f_\gamma + m_i f_\alpha$. It turns out that for a fixed value of n_i, the relative power in different m_i harmonics depends on the intrinsic parameters only through the inclination of the orbit, λ, in addition to κ. Using this information, we were able to estimate κ for the various sources and improve our estimate of λ. For all the challenges, these estimates gave similar values of κ ($\kappa \sim 1$). This lack of variation might indicate that κ is not being well determined - this will be investigated further in the future.

The second step, newly developed for this round of MLDC, used an implementation of the Metropolis-Hastings Monte Carlo (MHMC) algorithm to improve the parameter estimates. This MHMC search carries out matched filtering in the time-frequency domain, using as input the power along the tracks detected by CATS, and employing a theoretical template model for the time-frequency decomposition of an EMRI source based on the harmonics expansion described above. We use one chain consisting of 100000 points for each MHMC run. The priors for this chain were based on the challenge definition, except for the prior on plunge time, which was reduced to a narrower range based on the rough parameter estimates. The MHMC searches were initially seeded with the rough parameter estimates. In general we found that the chain moved away from the initial point, but not too far.

III. RESULTS

A. Challenge 1B.3.1

The combination of high SMBH mass and low eccentricity meant that there were a number of near-degeneracies in this data challenge, which made individual parameters difficult to resolve despite the high SNR. In particular, the MHMC search suggested that the strongest detected track for the $n = 2$ harmonic was the $m = 2$ sideband, and the secondary track was the $m = 1$ sideband. This corresponded to values of λ and κ that would have made the $m = -2$ sideband brighter than the $m = 1$ sideband, however. Since we could not detect another sideband in the presumed location of $m = -2$, we manually changed the inclination angle κ, which decouples from other parameters, to get a fit to the detected tracks.

B. Challenge 1B.3.2

The MHMC search pushed the SMBH mass all the way to the maximum value allowed by the prior, possibly indicating a lack of accuracy. The initial rough parameter estimate matched the data well, although not quite as well as the final MHMC result. We quoted both results, although this is artificial in a sense, because we have used the tight prior information to limit the movement of the MCMC chain, and then to decide that the MHMC output was not wholly accurate. If the chain had not been constrained, the best fit would have been for a slightly higher central black hole mass, around $5.26 \times 10^6 M_\odot$.

C. Challenge 1B.3.3

Again, the MHMC search pushed the SMBH mass to the maximum value allowed by the prior. This was, however, very close to the value detected by the initial rough estimate. Once again, if a less restricted prior was used, the best fit would have been for a slightly higher central black hole mass.

D. Challenge 1B.3.4

Although sections of multiple harmonics and sidebands were detected, the duration of each detectable segment was quite short. This suggested that parameter estimation might not be as accurate as usual. However, the final parameters reproduce the detected tracks well, which gave us some confidence in the results.

E. Challenge 1B.3.5

The tracks were quite difficult to find because of the low SNR. Nevertheless, we succeeded in finding a track corresponding to the $n = 3$ harmonic with a possible sideband near $t = 0$, and tracks corresponding to the $n = 2$ and $n = 3$ harmonics, each with one sideband, near $t = 3.6 \times 10^7$ s. The short duration of the tracks made accurate parameter estimation difficult, and the best fit parameters did not yield tracks that passed through the tracks detected near $t = 0$ (although they were close). We expected these parameters to be the worst of the five Challenges.

[1] Gair J R, Mandel I and Wen L, 2007, arXiv:0710.5250
[2] Gair J R and Jones G, 2006, *Class. Quantum Grav.* **27** 1145
[3] Cutler C, 1998, *Phys. Rev. D***57** 7089
[4] Wen L, Chen Y and Gair J R, 2006, Laser Interferometer Space Antenna: 6th International LISA Symposium, 873, 595