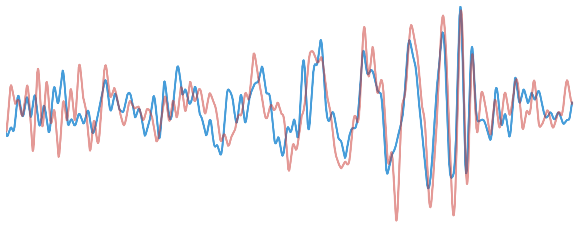
UNIVERSITY OF CAMBRIDGE



MPhil in Data Intensive Science

Submission: 23:59 on 22nd June 2025

GW minor module - Coursework Assignment

M. Agathos, J. Alvey, P. Canizares, C. J. Moore, and U. Sperhake

The coursework consists of three Parts: 1, 2, and 3.

The coursework will be submitted via a GitLab repository which will be created for you. You should write a report of no more than 3000 words to accompany the software you write to solve the problem. The approximate number of words that should be dedicated to explaining each part of the problem is indicated in square brackets; e.g. [x]. You should place all your code and your report in this repository. The report should be in PDF format in a " `report` *" folder. You will be provided access to the repository until the deadline above. After this, you will lose access, which will constitute the submission of your work.*

The software associated with the coursework should be written in Python and follow best software development practice as defined by the Research Computing module. It should be possible for someone else to easily run your code to reproduce all of the results included in your report.

Cosmology with Gravitational Waves

Gravitational wave (GW) signals propagate through the universe unimpeded by gas or dust. The signal amplitude encodes the luminosity distance, $D_{\rm L}$, to the source. They act as *standard candles*, or *standard sirens*, providing useful cosmological information.

GW sources are presumably located in galaxies. If the direction to the source can be measured from its GW signal, then it is possible to search in galaxy catalogs for likely host galaxies. The redshifts, z, of galaxies in the catalog are known from spectroscopy. These two pieces of information, $D_{\rm L}$ and z, enable a determination of the Hubble constant, H_0. This coursework problem asks you to do this for a simulated GW signal.

The following data files have been published along with this coursework:

- `H1_strain.txt`, `L1_strain.txt`, and `V1_strain.txt`. The measured frequency-domain strain data in the Hanford, Livingston and Virgo interferometers.
  First column: frequency [Hz]. Second: $\mathrm{Re}\,\tilde{h}(f)$ [s]. Third: $\mathrm{Im}\,\tilde{h}(f)$ [s].

- `H1_psd.txt`, `L1_psd.txt`, and `V1_psd.txt`. The noise power spectral densities (PSDs) in the Hanford, Livingston, and Virgo interferometers.
  First column: frequency [Hz]. Second: PSD [Hz^{-1}].

- `waveform.txt` Frequency-domain gravitational waveform up to luminosity distance correction. The waveform is generated at a fiducial distance of $D_{\rm L} = 1\,\mathrm{Gpc}$.
  First column: frequency [Hz]. Second: $\mathrm{Re}\,\tilde{h}_{+}(f)$ [s]. Third: $\mathrm{Im}\,\tilde{h}_{+}(f)$ [s]. Fourth: $\mathrm{Re}\,\tilde{h}_{\times}(f)$ [s]. Fifth: $\mathrm{Re}\,\tilde{h}_{\times}(f)$ [s].

- `mock_galaxy_catalog.txt`. A list of host galaxies for Part 3 of the project.
  First column: name. Second: right ascension [rad], Third: declination [rad]. Fourth: redshift.

Part 1: Matched filtering time delays

For this part of the question, you may assume that the GW signal is known and is given in the file `waveform.txt`. You may also adapt the code from the first examples class to answer this question.

1. Using the measured strain data in the H1 and L1 interferometers, perform a separate matched filter search in each detector using the GW waveform and PSD information provided. You should plot the matched filter signal-to-noise ratio (SNR) as a function of time in each detector. Using the peak of these SNR curves, determine the time at which the source coalesces in each detector: $t_c^{(\mathrm{H1})}$ and $t_c^{(\mathrm{L1})}$. Using the full width half maximum of the peak in the SNR curves, estimate the uncertainty on these times. [*Info: All strain data provided covers a data segment starting at*

GPS time of 1126259460.4 GPS s *with a duration of* 4 s *and sampling frequency* 2048 Hz. *In this question and below, you may assume that the antenna patterns of the detectors do not vary significantly across the time segment considered. You may use the detector antenna functions provided in* `bilby` *via the* `antenna_response` *method in the* `Interferometer` *class. You should* ***not*** *use any in-built code for computing the matched filter.*] [400]

2. Using the measured time delay $t_c^{(\mathrm{H1})} - t_c^{(\mathrm{L1})}$ and the known locations of the L1 and H1 interferometers, what can be said about the direction to the GW source? Produce a plot of the sphere of the sky (i.e. a *sky map*) showing the area consistent with your measurement of the delay. You may use any suitable map projection of your choice. Describe and explain the key features of your sky map. [*You may use the* `time_delay_from_geocenter` *method in the* `Interferometer` *class in this question.*] [400]

3. Load the measured strain data for the V1 interferometer. Perform the matched filter search of this data to determine $t_c^{(\mathrm{V1})}$. Explain why it makes sense to consider the Virgo interferometer separately from the other two. [100]

4. Produce an updated sky map showing the area consistent with your measurement of the three time delays $t_c^{(\mathrm{L1})} - t_c^{(\mathrm{H1})}$, $t_c^{(\mathrm{H1})} - t_c^{(\mathrm{V1})}$, and $t_c^{(\mathrm{V1})} - t_c^{(\mathrm{L1})}$. Even though Virgo may be less sensitive than the two LIGO interferometers, including it in the analysis can have a large effect on the determination of the sky area; estimate the factor by which the sky localisation is improved when including V1. [400]

Part 2: Bayesian inference

For this part of the question, you may again assume that the GW signal is known and is given in the file `waveform.txt`. The parameters that need to be inferred are D_L, `ra`, `dec`, ψ, $t_c^{(\mathrm{geo})}$, where the final parameter is the coalescence time at the Earth geocenter.

1. Write the likelihood $\mathcal{L}(\mathrm{data}_{\mathrm{IFO1}}, \mathrm{data}_{\mathrm{IFO2}}, \ldots | D_\mathrm{L}, \mathtt{ra}, \mathtt{dec}, \psi, t_c^{(\mathrm{geo})})$. You may use the `bilby` package for the locations of the interferometers on Earth, time delays between the detectors and the geocenter, and the antenna patterns F_+ and $F_\times$, but you should write your log-likelihood function. [400]

2. Using the L1 and H1 data, use a stochastic sampling package of your choice to sample the 5-dimensional posterior on the parameters listed above. Describe the priors you use and briefly explain your choices. [*A standard choice in GW data analysis is the* `dynesty` *package, which is a nested sampling package provided with* `bilby`*. If you choose to use this sampler, it may be useful to investigate the impact of the* `bound` *and* `sample` *options in the* `NestedSampler` *class on runtimes. You should* ***not*** *use a pre-implemented version of the likelihood* $\mathcal{L}$.] [400]

(TURN OVER

3. Plot the posterior on the angles `ra` and dec as a sky map. Compare this posterior with your answer from Part 1.2. Plot the 1-dimensional marginalised posterior on the luminosity distance, D_{L}. [200]

4. Redo the Bayesian inference including the data from all three detectors: L1, H1, and V1. Produce an new sky map showing the 3-detector posterior. Compare this posterior with your answer from Part 1.4. Plot the new 3-detector posterior on D_{L}. Comment of the effect of including Virgo in the analysis on the sky localisation map. [200]

Part 3: GW cosmology

1. Look at the galaxy catalog file provided. Using the posterior computed from Part 2.2 (i.e. only using the L1 and H1 data), produce a list of all galaxies contained in the 90% confidence interval on `ra` and `dec`. If you now include information from V1, what is the name of the most probable host galaxy? [*Provided that you give sufficient evidence for your method and results, Part 3 will be marked using your generated posterior samples from Part 2.*] [250]

2. Assuming that the most probable galaxy found in the previous part of the question is indeed the host, measure the Hubble constant H_0 accounting for the uncertainty in the luminosity distance D_{L}. [*You may assume that the redshift of the host is perfectly measured.*] [250]