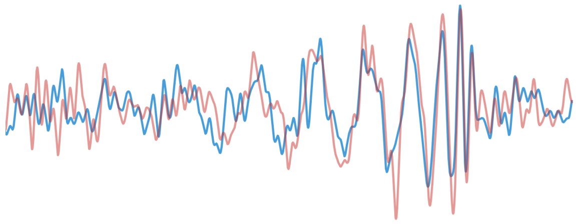
UNIVERSITY OF CAMBRIDGE



MPhil in Data Intensive Science

Submission: 11:59pm on Friday the 21st of June

**Coursework Assignment - Gravitational Waves minor module**

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

*The coursework consists of three parts: 1, 2 and 3. Equal marks are available for each part.*

*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. 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.*

*You are expected to submit code and associated material that demonstrates good software development practices as covered in the Research Computing module.*

*Your report should not exceed 3000 words (including tables, figure captions and appendices but excluding references); please indicate its word count on its front cover.*

*You are reminded to comply with the requirements given in the Course Handbook regarding the use of, and declaration of, autogeneration tools for both the code and report.*

## Part 1: Waveform Model

In this part of the coursework you will work with the waveform model `IMRPhenomD`. This is a model for the GW signal from the inspiral-merger-ringdown of a binary black hole.

The waveform model is implemented in the `pycbc.waveform` library and can be accessed using the function `get_fd_waveform`.

For this part of the question you should use the following parameters for the binary black hole: primary mass $m_1 = 30\,M_\odot$, secondary mass $m_2 = 30\,M_\odot$, distance $d = 500\,\mathrm{Mpc}$. You should generate you waveform using a frequency spacing of $\delta f = (16\,\mathrm{s})^{-1}$ and starting from a lower frequency cutoff of $f_{\mathrm{lower}} = 10\,\mathrm{Hz}$.

1. Use the `get_fd_waveform` function to produce the frequency-domain waveform model for the source with the parameters described above when viewed from an inclination angle $\iota = \pi/3$ and a phase angle of $\Phi = 0$. The function returns polarisation states $\tilde{h}_+(f)$ and $\tilde{h}_\times(f)$. Produce a log-log of the waveform amplitude $|\tilde{h}_+(f)|$ as a function of frequency. From this plot, estimate the approximate frequency at which the merger of the two black holes occurs.

2. Consider a single GW interferometer detector with an expected noise amplitude spectral density (ASD) $\sqrt{S(f)}$ given in the file `ASD.txt`. Produce a log-log of this ASD as a function of frequency.

3. On the same set of axes, produce a log-log plot of both $\sqrt{f}|\tilde{h}_+(f)|$ and $\sqrt{S(f)}$ as a function of frequency. State clearly the dimensions of all the quantities in this plot.

4. Define the antenna pattern functions $F_+(\theta, \phi, \psi)$ and $F_\times(\theta, \phi, \psi)$ for an $L$-shaped interferometer. Explain how these antenna patterns can be used to relate the waveform polarisations $\tilde{h}_+(f)$ and $\tilde{h}_\times(f)$ to the measured detector output $\tilde{h}(f)$.

5. For a detector with this ASD, compute the signal-to-noise ratio (SNR) for an optimally orientated source with the parameters given above. I.e. for the sky position angle parameters $\theta$, $\phi$, $\psi$, and the source orientation angles $\iota$ and $\Phi$ you should choose the values that give the largest possible SNR. (Use $f_{\mathrm{upper}} = 1000\,\mathrm{Hz}$.)

$$(\text{optimum SNR})^2 = 4\int_{f_{\mathrm{lower}}}^{f_{\mathrm{upper}}} \mathrm{d}f\, \frac{|\tilde{h}(f)|^2}{S(f)}$$

6. Compute the angle-average optimum SNR for the source with the parameters given above. I.e. instead of maximising over all the angles you should now take an average over them using a physically appropriate probability density function (PDF). State clear what PDF you use for each angle.

## Part 2: Matched Filtering

In this part of the question you will use the waveform model from part 1 to perform a search for a signal in a long stretch of noisy data using a matched filter.

Matched filtering can be shown to be the optimal method for detecting known signals in Gaussian noise. This is the method that allows us to detect GW signals using our detectors even when the amplitude of the GW is much less than the amplitude of the noise.

The measured detector data is provided in the file named `data.txt`. This file contains two columns: 1) sample times [seconds] and 2) measured strain [dimensionless].

1. Download the simulated data time series $d(t)$ from the Moodle page. Plot the data as a function of time.

2. Using a Welch periodogram, determine the measured PSD $S_n(f)$ for this data in the frequency range $10 < f/\mathrm{Hz} < 10^3$. Hence determine the measured ASD. Plot the measured ASD as a function of frequency using a log-log axis scaling. Does the measured ASD agree with the expected ASD used in the first part of the question?

3. Using the PSD found in the previous part of the question, "whiten" the data. Plot the "whitened" data as a function of time. Can you see the signal in the whitened data?

4. Using the `IMRPhenomD` waveform model from the first part of the question, calculated the quantity $\tilde{h}_+(f)$ for a binary with parameters $m_1 = 30\,M_\odot$, $m_2 = 30\,M_\odot$, $\iota = \pi/3$, $\Phi = 0$. This will be your template signal. Scale the amplitude of your template such that it has an optimum SNR of 1. Using this (scaled template) search search the data for a signal. Plot the matched filter SNR as a function of time. The matched filter SNR is defined as

$$(\text{matched filter SNR})^2 = 4\left|\int_{f_\text{lower}}^{f_\text{upper}} \mathrm{d}f\ \frac{\exp(2\pi i f t)\tilde{h}(f)\tilde{d}^*(f)}{S(f)}\right|$$

5. From the plot of matched filter SNR versus time. Identify the time of the merger in the data.



(TURN OVER

## Part 3: Parameter Estimation

In this part of the coursework you will use the waveform model from part 1 to perform Bayesian data analysis on a short stretch of data where you have presumably identified the presence of a gravitational wave signal.

We have seen how Bayesian statistical inference algorithms, such as MCMC and nested sampling are used for sampling the posterior distribution on our model's parameter space, given a noise PSD.

Here, you should use `dynesty`, a nested sampling package in Python, and will equip it with some additional functionality to perform parameter estimation on GW data, with minimal dependencies only on standard python libraries. The `dynesty` package implements the nested sampling algorithm, including some modern features for increased efficiency. You will find the complete documentation for `dynesty` at https://dynesty.readthedocs.io. To adapt it to the specifics of our GW data analysis problem, we need to provide the `dynesty` sampler with i. a *likelihood* function $L(d|\boldsymbol{\theta})$, and ii. a *prior* $\pi(\boldsymbol{\theta})$ on our parameter space. Implicitly, the likelihood function will make use of iii. a *model* $h(f;\boldsymbol{\theta})$ that for any point on the parameter space provides a waveform in the frequency domain.

1. Define the parameter space $\boldsymbol{\Theta}$ for the IMRPhenomD model. Write down the list of its component parameters $\theta_i$, their dimensionality and units, and their topology.
2. Write a wrapper function `waveform_model` that takes a list of BBH parameters as arguments, together with a frequency array with the grid values $f_i$, and returns two frequency series for the two polarisations $h_+(f_i)$ and $h_\times(f_i)$ based on the *IMRPhenomA* model that you coded up in Part 1.
3. Code up a *likelihood* function $L(d|\boldsymbol{\theta})$ in the frequency domain, based on the hypothesis of zero-mean stationary Gaussian noise for a single interferometric detector.
4. Write a `prior transform` function that is compatible with the `dynesty` sampler. (You should choose your own physically reasonable priors for each parameter. You should justify your choices in the report.)
5. Run a Bayesian analysis on a short, 8 second long segment of data centered on where you think the signal is located based on the results of your analysis from part 2.
6. Plot the posterior probability distributions.