GRIII 2021 Capstone Project

For your capstone project you will be working with actual LIGO data available at the Gravitational Wave Open Science Website `https://www.gw-openscience.org/about/`. In addition to the data, you will also find many helpful tutorials and codes at this site. Note that the hyperlinks in this document should all be clickable, taking you right to the relevant page.

The goal is to analyze the data surround the first observation of a binary black hole, GW15094. You should study the discovery paper `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.116.061102`. Much of what you need to know for this project can be found in the LIGO/Virgo data guide paper (which I wrote large sections of) `https://iopscience.iop.org/article/10.1088/1361-6382/ab685e`.

This project will involve writing code. I don't care what language you use, so long as it is C or Python. If you haven't coded before it is time to learn. But don't worry, most of the code you need has already been written, especially if you choose to work in Python. The GWOSC tutorials include many examples of Python code for reading LIGO data and performing analyses. The most relevant code is the Python companion to the data guide paper: `https://github.com/losc-tutorial/Data_Guide`. You can also view the Jupyter notebook `https://colab.research.google.com/github/losc-tutorial/Data_Guide/blob/master/Guide_Notebook.ipynb`. If you prefer to work in C, the codes you need are at `https://colab.research.google.com/github/losc-tutorial/Data_Guide/blob/master/Guide_Notebook.ipynb`.

Finding the Signal

The goal here is to analyze the LIGO Hanford and LIGO Livingston data and find the GW150914 signal. To cut down on the computational cost we will focus our attention on the 4 seconds of data shown in Figure 3 of the Data Guide paper. You can use the downsampled 4096 Hz data. To further simplify things, you can ignore the spin of the black holes and focus on systems with component masses between $25M_\odot$ and $45M_\odot$. You can use the very basic IMR-PhenomA waveform model `https://journals.aps.org/prd/abstract/10.1103/PhysRevD.77.104017`, which has been coded up by my former student Travis Robson here `https://github.com/eXtremeGravityInstitute/LISA_Sensitivity/blob/master/PhenomA.py`. The code uses the total mass $M = m_1 + m_2$ and the symmetric mass ratio $\eta = m_1 m_2/M^2$. I want you to work with the component masses, so you will need to compute the total mass and symmetric mass ratio to feed to the subroutine. The quantities you need are the frequency domain amplitude and phase (called Aeff and Psieff in the code), from which you can construct the frequency domain waveform $h(f) = A(f)e^{i\Psi(f)}$.

The main quantity you need to calculate is the SNR time series $\rho(t, \boldsymbol{\mu}) = |z(t, \boldsymbol{\mu})|$, described in my notes and in equations (15) through (21) of the data guide. It can be computed cheaply using the inverse discrete Fourier transform of equation (21). You can analyze the Hanford and Livingston data separately, so there is no need to worry about antenna patterns and time delays. These simply rescale the amplitude, shift the phase and shift the merger time. Put another way, the amplitude, phase and merger time you compute are in the detector frame. The relative amplitude, phase and arrival time between the detectors encode the information about the sky location, inclination and polarization (see equations 25 and 26 of the guide), but we are not worrying about those for this exercise. The SNR time series is automatically maximized over the phase offset and amplitude. The remaining *intrinsic* parameters $\boldsymbol{\mu}$ are the component masses m_1, m_2. For each set of parameters $\boldsymbol{\mu} = \{m_1, m_2\}$ you get an SNR time series, and the maximum of each time series can be read off using a simple sort.

A key quantity that you need to be able to estimate is the power spectral density (PSD) $S_n(f)$ for each detector which appears in the noise-weighted inner products. The Python notebook for the Data Guide paper takes you though the process of computing the spectral estimates using Welch averaging (after first applying a Tukey filter to prevent spectral leakage). You can use the supplied code to compute the PSDs.

To conduct a LIGO-style search you will need to lay out a template grid in m_1, m_2. Before getting to that, compute the SNR time series for a signal with detector frame masses $m_1 = 39M_\odot$ and $m_2 = 32M_\odot$. You should find a maximum SNR of about 19 in Hanford and 13.5 in Livingston. Make a plot of the SNR time series in each detector. You should find that the Livingston time series peaks about 7 ms before the Hanford time series. What do you find for the phase difference? Try a grid of masses between $5M_\odot$ and $100M_\odot$ spaced by $5M_\odot$ or less and plot the quadrature combined maximum $\rho = (\rho^2_{\rm max,H} + \rho^2_{\rm max,L})^{1/2}$ in the two detectors. Where does it peak? If you manage to get this far into the project you are doing well!

To lay out a search grid we need to compute the overlap, or match between signals with parameters $\boldsymbol{\lambda}, \boldsymbol{\theta}$:

$$\mathrm{M}(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \frac{(\mathbf{h}(\boldsymbol{\lambda})|\mathbf{h}(\boldsymbol{\theta}))}{\sqrt{(\mathbf{h}(\boldsymbol{\lambda})|\mathbf{h}(\boldsymbol{\lambda})(\mathbf{h}(\boldsymbol{\theta})|\mathbf{h}(\boldsymbol{\theta})}}$$

For nearby signals we can write $\boldsymbol{\theta} = \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}$ and Taylor expand in $\Delta\lambda^\mu$:

$$\mathrm{M}(\boldsymbol{\lambda}, \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}) = 1 - \frac{1}{2}\left(\frac{(h_{,\mu}|h_{,\nu})}{(h|h)} - \frac{(h|h_{,\mu})(h|h_{,\nu})}{(h|h)^2}\right)\Delta\lambda^\mu\Delta\lambda^\nu + \dots$$

The quantity in brackets is called the template metric $g_{\mu\nu}(\boldsymbol{\lambda})$. Here the parameters are $\boldsymbol{\lambda} \to \{m_1, m_2, t_0\}$ (The ϕ_0 parameter has already been removed by the maximization). Since we maximize over t_0, we can project it out via

$$g'_{\mu\nu} = g_{\mu\nu} - \frac{g_{\mu t_0} g_{\nu t_0}}{g_{t_0 t_0}}$$

The only non-zero components of the metric $g'_{\mu\nu}$ are for $\{\mu, \nu\} = \{m_1, m_2\}$. The metric is $g'_{\mu\nu}$ not constant - it depends on m_1, m_2. Compute the metric at $m_1 = 39M_\odot$ and $m_2 = 32M_\odot$, and ignoring the fact that it is not constant, find the spacing for a rectangular grid in m_1, m_2 that would ensure matches $M \geq 0.97$ between any signal and the templates in the grid. Using this grid spacing, search the triangular mass range $m_1 \in \{25M_\odot, 45M_\odot\}$ with $m_1 > m_2$. Find the maximum SNR at each grid point and plot the result as a heat map for each detector. For what masses is the SNR maximized in each detector?

For bonus points, you could try and write a simple MCMC to explore the four dimensional parameter space $\boldsymbol{\lambda} \rightarrow \{m_1, m_2, t_0, \phi_0\}$ using the un-maximized log likelihood $\ln p(d|\boldsymbol{\lambda}) = -\frac{1}{2}(d-h|d-h) + \mathrm{const.}$ with uniform priors on the four parameters. The full Fisher information matrix $\Gamma_{ij} = (h_{,i}|h_{,j})$ could be used as a proposal, as described in my Saas Fee lectures. You should submit a write up and your codes.