Today, gravitational waves are one of the instruments to study astrophysics systems and to test general relativity. They are detected through interferometers, such as LIGO and VIRGO. These are powerful and sensitive detectors since GW signals are relatively small. For these reasons, they are extremely sensitive to noise sources, such as seismic oscillations, instrumentation vibrational resonances and so on. Since the output of the detectors are time signals, it's dominated by noise fluctuations. We must find a way to extract the GW signal from the chaos.

Fortunately, this is a well known problem in physics and in general in signal analysis. The ground basis from we must start is the very nature of our data stream. Our signal is discrete, since we are able to sample the source only every a certain amount of time. At the same time, the collection of our points are influenced by an intrinsic stochastic noise. Nevertheless, we have a fair amount of expected signal templates to search for, but they depend on physical parameters, which we want to measure.

These features allow us to treat our data stream as a stochastic process, or time series in this particular case. Basically, our collection of sample, the realization, is just one observation of a hidden function. Sample have a certain probability distribution.

Every process is characterized by the mean function of time, which is the expected value of y_i over the realizations ensamble, and the autocovariance matrix, whose definition is written here, which stores information about the independence of observing two points i and j. We could normalize the autocovariance matrix to obtain the autocorrelation matrix.

There are some special kinds of process, interesting for our problem. One of these are stationary processes. A process is stationary if mean is constant in time and autocovariance matrix is proportional to identity. Another type of processes are the so-called wide sense stationary processes. In these cases, mean is again constant in time, but we relax the constraint on autocovariance: we just need that the autocovariance matrix depends only on the difference of time of two samples.

To test if our process belongs to one of these categories, we start with the mean function evaluation. To see the time evolution, we simply average contiguous 5-seconds chunks. We see that the mean is constant within 2 standard deviations and the total final average is compatible within zero. So in the next steps we will forget about mean and autocovariance and autocorrelation are the same.

The next step is about the variance of the process. We start evaluating the autocorrelation function (ACF) through the formula showed. ACF performs a 2 points correlation in function of the time-lag between the points. We see that, due to our normalization, the 0 lag ACF is 1, since we are evaluating the correlation between our signal and itself. At the next step we get the correlation between signal and its 1-lag translated copy and so on.

ACF is the right instrument to test stationarity: if we zoom on the first lags, we clearly see that ACF is different from 0 for non-zero lags. This is sufficient to break the stationarity hypothesis. But ACF gives us other information, since implicitly defines a timescale over which the process is stationary. This quantity is called ACTime, and it is the timelag of the first zero of the ACF. Our 1% measure is showed.

This means that every ACT, the process is stationary. To better see this, I plotted the ACF computed over the samples which correspond to an integer multiple of ACT. We see that ACF is 0 within fluctuations on every point beyond zero-lag. This operation is called thinning

Since we need the entire sample, this solution is not enough though. We should test WSS hypothesis. In order to do this, we note that in the WSS hypothesis, autocorrelation depends only on time intervals. This means that the ACF computed every dt should be invariant. So we split our sample in 1-seconds chunks and compute ACF. For the sake of visualization, in the figure we compare 3 ACFs.

By the way, we need a quantitative esteem of how much ACF changes. The solution adopted is ACT. The plot show the values of ACT computed for each chunks. We see that the majority of values are spread over the same value. With some grade of approximation, our process is WSS.

Now that we know everything about our process, we can think about the extraction. Since we are using statistical inference, we need a way to assign likelihood to our problem. We assume that our data are the result of the superposition of noise and signal. In order to compute our data distribution, we marginalize over the unobserved features. Using product rules and integrating over both signals points and noise we get that the likelihood function is the noise distribution evaluated in the residuals. We simply translated our ignorance into the problem of finding a suitable noise distribution

To summarize, all we know about the process is in the autocovariance matrix, since its mean is 0. A way to resolve the problem is using the max information

entropy principle. The probability distribution which best represents the current state of knowledge about
a system is the one with largest entropy, the Shannon's functional. Our state of knowledge is translated as a constraint in the maximization routine. The result of the calculation under known covariance matrix leads to a multivariate gaussian distribution, which gives us our noise distribution

We should use the WSS process advantages. Autocorrelation matrix in this case is function only of ACF at a given time interval. The symmetric matrix belongs to the Toeplitz class, which, for large samples, asymptotically converge to a circulant matrix. Circulant matrices are well studied and can be shown that the diagonalization process consists in a discrete fourier transform.

Diagonal form contains the so called power spectral density of the process, which can be thought as the variance of the noise at any given frequency. In fact, PSD defines completely the nature of the noise of our process. This is a consequence of a much general result, the Wiener Khinchin theorem. At the end of this procedure we have our simpler likelihood, at the cost of working in the frequency domain.

We compute the PSD by a real fast fourier transform algorithm of ACF computed over the entire data sample. In the figure we see the spectrum focused on the physical interesting frequencies. Since the spectrum has evident peaks in some frequencies, we should talk about a coloured process, different from white processes in which all of the available frequencies contributes in the same manner to the noise spectrum.

We can use PSD to make another test on WSS. We evaluate PSD on 8-seconds chunks of the global sample. In this figure are reported three PSDs and we can see that they are pretty superimposed.

I wanted to try also another way to compute PSD. I saw Welch's method in literature. This methods consists in computing FFT in overlapping chunks using a specific window, then average over the chunks. This reduce variance of PSD but also resolution. Nevertheless I wanted to try and I think that PSD is more readable in this way. For the calculation I used 8-seconds chunks with 4 seconds overlap and a Tukey window.

The last step is to extract the signal. Since our template depends on parameters, we must find optimal values. In order to do this, we could maximize the likelihood function and then to get the most likely template. But instead we are in the conditions where we should apply a filter to our data to maximize likelihood.

This filter is called matched filter, or Wiener filter. We show that likelihood function is writable as a inner product between residuals. The mixed term in the product is the searched filter. The connection is clearer if we define the signal-to-noise ratio. From this point of view, to max SNR correspond max likelihood.

So, in order to extract our signal we need a signal template. For this analysis we used a simplified TaylorF2 template to 2 order in post newtonian approximation. This template is a function of frequencies and depends on multiple parameters, such as the time of arrival of the GW and black hole masses. Mass dependence is parameterized by using chirp mass and mass ratio, whose definition are shown in this slide. For our purposes we neglected frequency and mass dependence in amplitude.

Since we want to infer black hole masses, to use Bayes theorem we need prior distributions, based on our current state of knowledge about the two black holes masses. Since we are using chirp mass and mass ratio, we must perform a change of variables to propagate our probability distributions. The presented formula allows us to easily change variables, where the *f* function is the transformation to single masses to the new paramteres. The determinant of the jacobian is fundamental in order to conserve probability through the transformation.

In this slide there are some sketches of the actual calculation. We have the inverse relation for the variable transformations, from which we compute its jacobian. If we compute its determinant, we are lucky enough to have a fairly simple expression. Given the values suggested in the assignment and assuming a uniform prior over this intervals, we obtain the joint probability of chirp mass and mass ratio, which is proportional to the determinant. In the plot we show the 3d visualization for the unnormalized distribution using a colormap.

Here are the two normalized marginalized priors.