Detection of transient signals in time series: application to gravitational wave data

Filippo Cattafesta

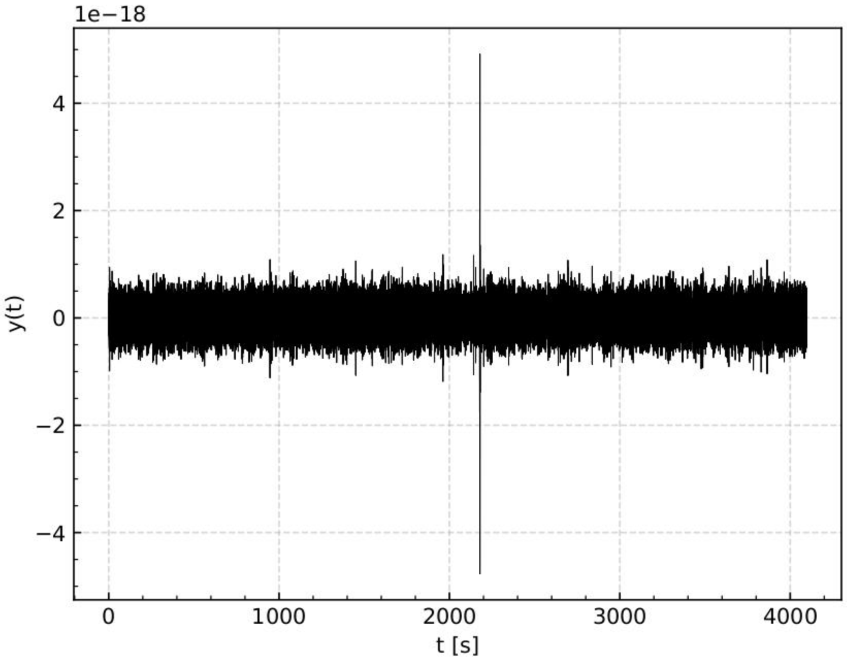
Introduction to Bayesian probability theory

GW signal extraction problem

- Interferometers are used to detect Gravitational Waves
- Extremely sensible to GW signals and to noise sources

At the end of the day, detectors output looks like this…

We need a way to detect the interesting signal



Our problem in detail

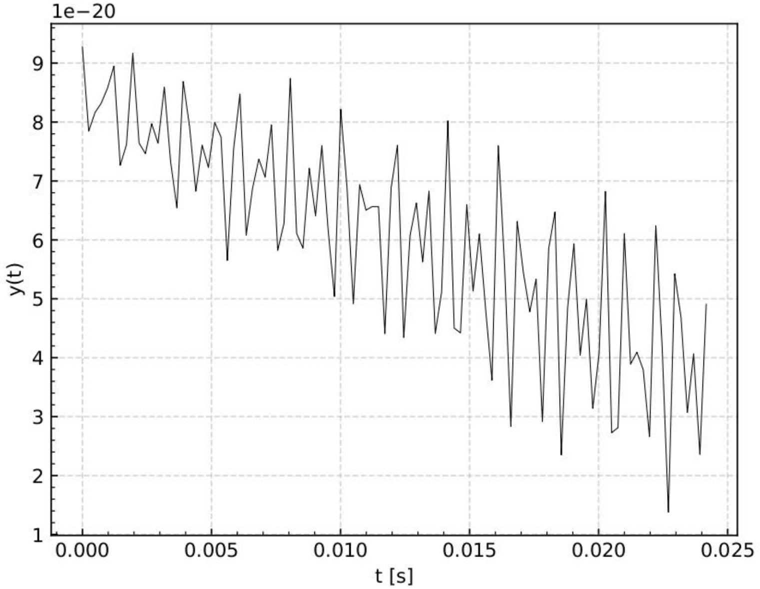
- The output is a discrete signal in time

$$(y_i, t_i)$$

result of an intrinsic stochastic unknown noise.

- We know what we are looking for (PN templates for GW signals) but we should be able to make measurements of interesting parameters.

We must treat our problem in a statistical way



Using stochastic processes

We observe a realization of a hidden function $f(t)$

$$(y_i, t_i)$$

where y_i have their own probability distribution

$$y_i \sim P(y_i | t_i \boldsymbol{\theta}).$$

- We are assigning to our discrete signal an unknown *pdf*
- Special features of the process could help us to extract our signal

Fundamental features of the process

- Mean function

$$\overline{y}(t) = E[y_i]$$

- Autocovariance matrix

$$C_{ij} = E[(y_i - \overline{y})(y_j - \overline{y})]$$

which tells us if the probability of observing y_i is independent of y_j.

Stationarity

- A process is stationary if

$$\overline{y}(t) = \text{const.} \qquad C_{ij} = \delta_{ij}\sigma^2$$

- A weaker requirement is

$$C_{ij} = C(|t_i - t_j|)$$

which defines the Wide Sense Stationarity.

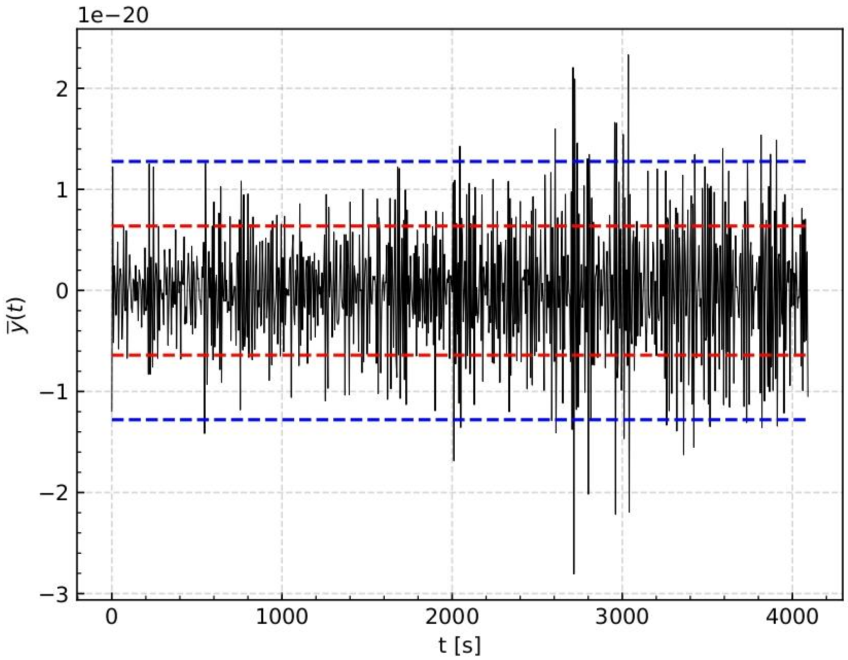
What about our process?

- Mean is computed getting the average of 5 seconds chunks

- Mean is constant within 2 standard deviation

- Mean all over the sample is

$$\overline{y} = (0 \pm 6) \times 10^{-21}$$



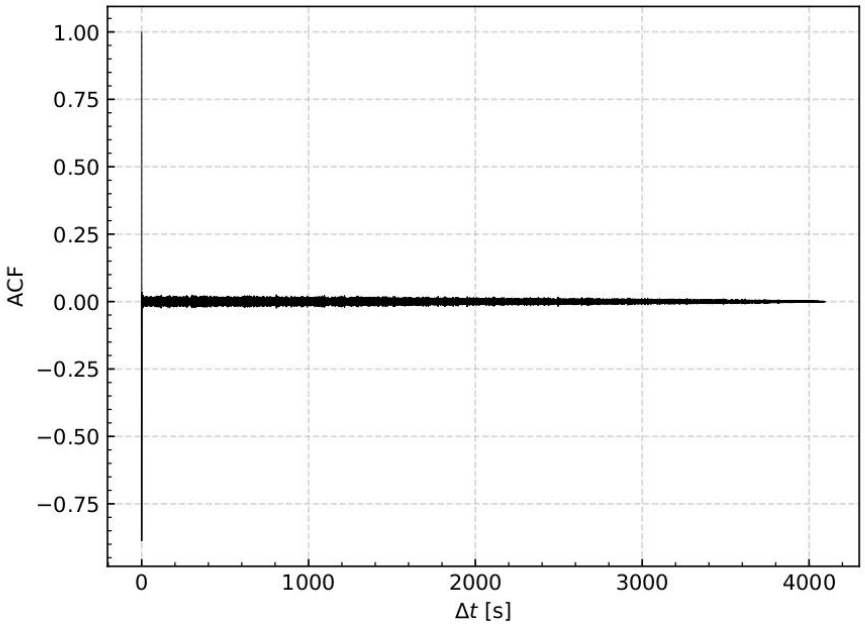
What about our process?

- In order to test stationarity we analyze the Autocorrelation Function

- We compute ACF using `scipy.signal.correlate`

$$ACF(k) = \frac{1}{n\mathrm{var}(\boldsymbol{y})} \sum_{l=0}^{n-1} y_l y_{l-k+n-1}$$

where $k = 0, 1, \ldots, 2n-2$



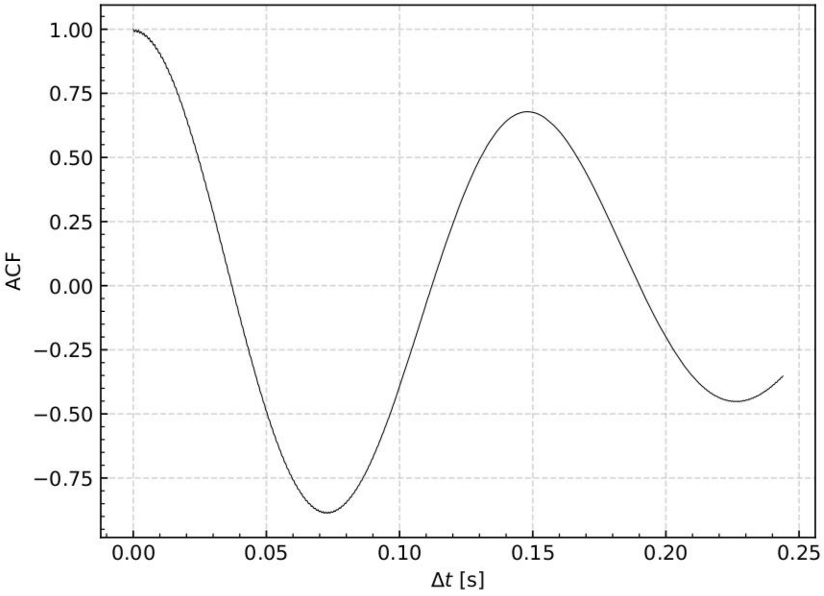
Stationarity

- Since ACF is non-zero, process is not stationary

- Nevertheless, there is a timescale over which the process is stationary

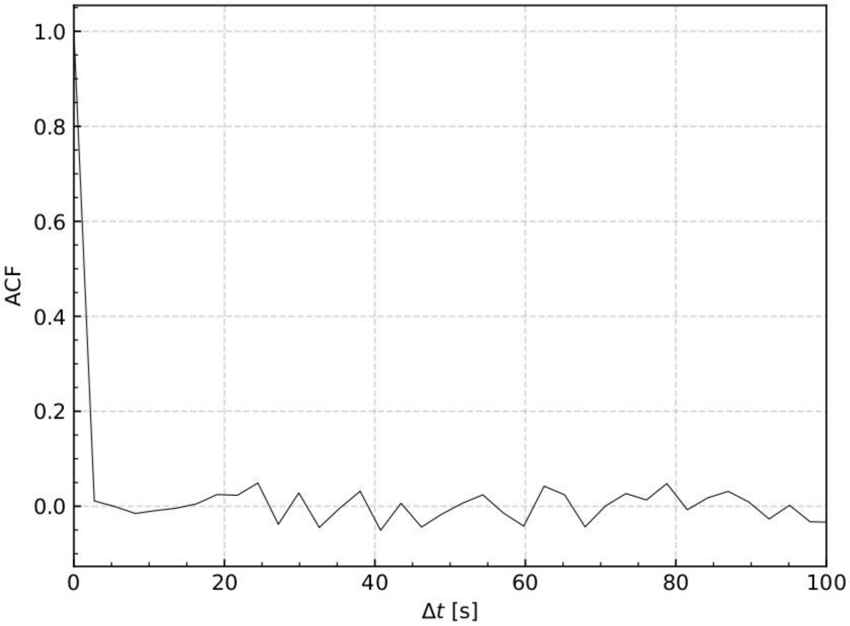
- This quantity is ACT, which is the lag time of the first ACF zero

$$ACT = 0.368 \pm 0.004 \text{ s}$$



Stationarity

- Thinning shows how ACT works
- ACF of a stationary process (within fluctuations)

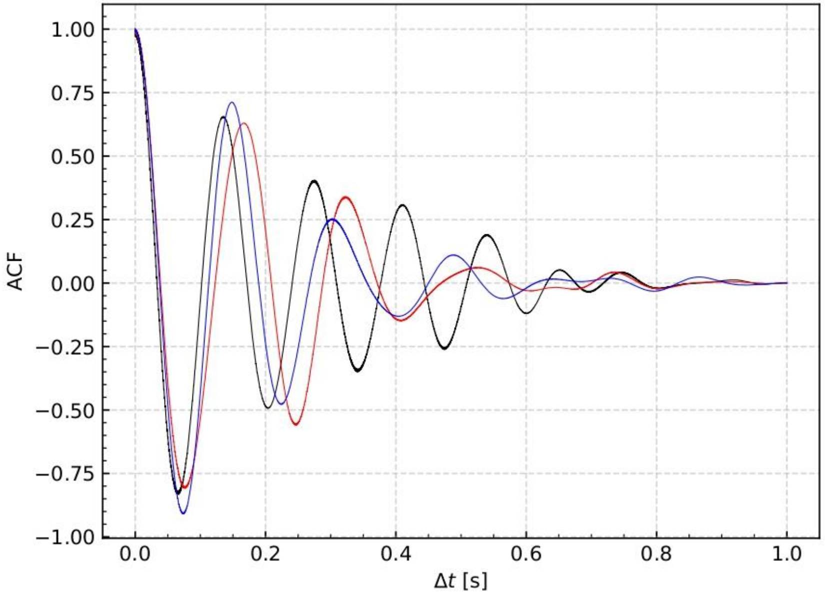


Wide sense?

- For a WSS process, autocorrelation depends only on the time interval

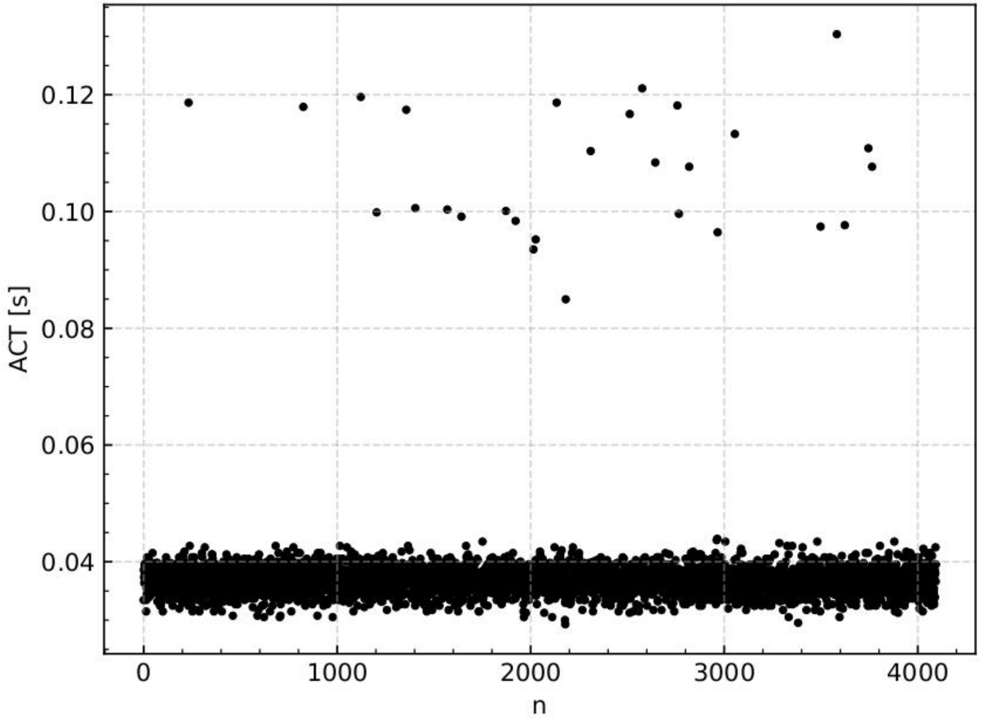
- We split the data sample into 1 seconds chunks and compute ACF for each interval

- ACF should be invariant for every interval



Wide sense?

- How can we measure how much ACF changes in each chunk?
- A solution could be ACT: the majority of ACT is spread over the same value
- Our process is compatible with WSS hypothesis



How to assign likelihood?

Under signal-noise assumption

$$d_i = n_i + h_i$$

likelihood is computed marginalizing the joint distribution over noise and predictions

$$P(d_i|HI) = \int dn_i \, dh_i P(d_i, h_i, n_i|HI)$$

which, for a deterministic model, leads to

$$P(d_i|h_i) = g(d_i - h_i)$$

where g is the noise distribution.

Max-entropy principle

- Since the process has $\overline{y} = 0$, our non trivial background information is all about autocorrelation matrix

- The probability distribution which best represents the current state of knowledge about a system is the one with largest entropy

$$H_n = -\sum_i^n p_i \log p_i$$

Maximization problem under known covariance matrix constraint leads to a multivariate gaussian distribution

$$P(n_i) = \exp\bigg(-\frac{n_i C_{ij}^{-1} n_j}{2}\bigg)$$

WSS solution

- Assuming WSS, autocovariance matrix depends only on ACF

$$C_{00} = ACF(0) \qquad C_{01} = C_{10} = ACF(\Delta t) \qquad \dots$$

- Toeplitz matrix with asymptotically circulant matrix behaviour

- We are able to diagonalize circulant matrices through Discrete Fourier Transform

$$\begin{pmatrix} ACF(0) & ACF(\Delta t) & \cdots & ACF(n\Delta t) \\ ACF(\Delta t) & ACF(0) & \cdots & ACF((n-1)\Delta t) \\ \vdots & & \ddots & \\ ACF(n\Delta t) & \cdots & \cdots & ACF(0) \end{pmatrix}$$

Likelihood in frequency domain

- Assuming WSS, we get the diagonal form of the autocovariance matrix

$$S_i = F(\mathrm{ACF}(i\Delta t))$$

and it is called Power Spectral Density (Wiener-Khinchin theroem).

- Finally, our likelihood takes the form

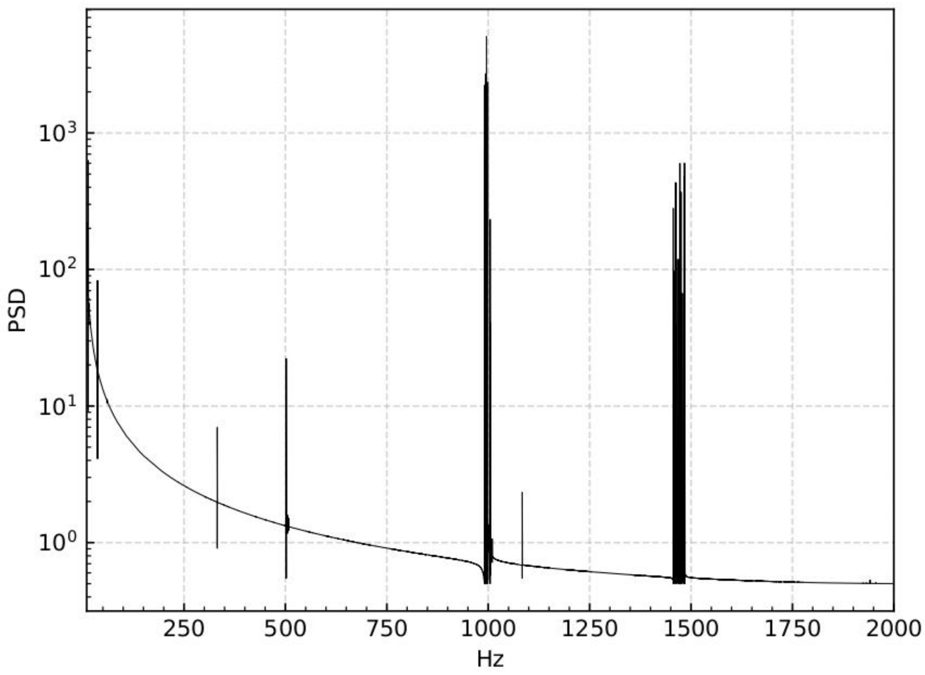
$$P(\widetilde{d}_i|\widetilde{h}_i) = \exp\left(- \frac{|\widetilde{d}_i - \widetilde{h}_i|^2}{2S_i} \right)$$

PSD of our process

- We take the Real Fast Fourier Transform of both data and ACF using `scipy.fft.rfft`

- Focusing on physically relevant frequencies

- Noise is clearly colored, since spectrum contains peaks at certain frequencies

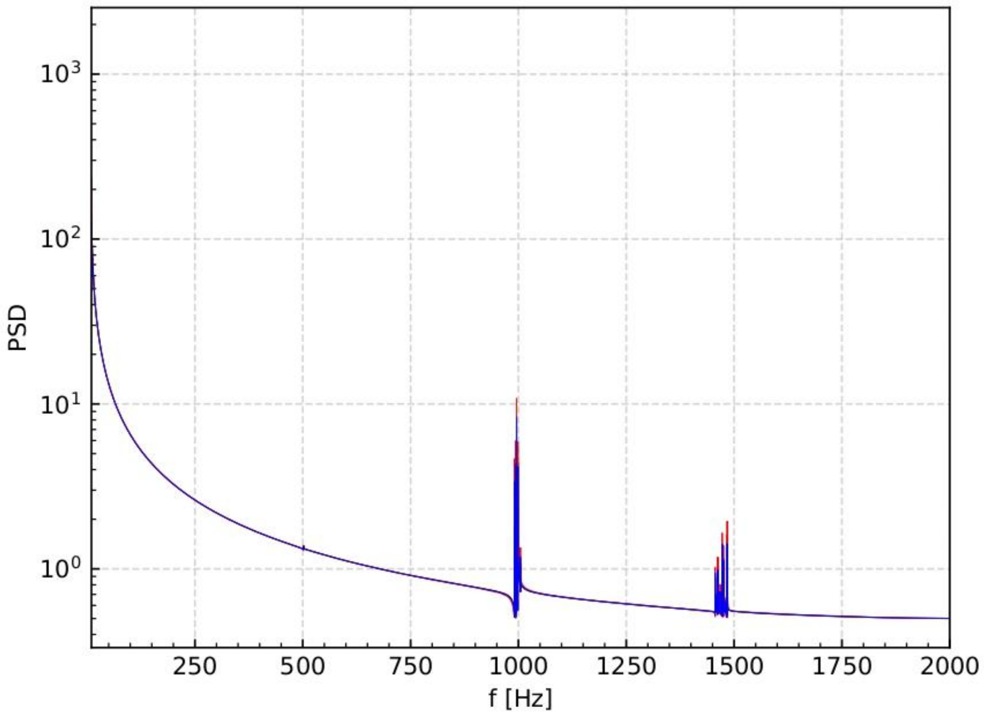


Another WSS test

- To test WSS we can analyze PSD too

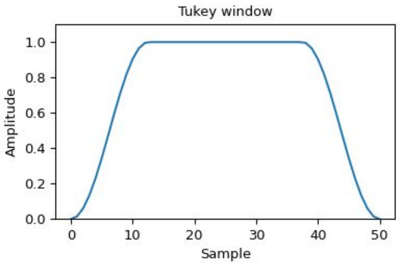
- We compute PSD of 8 seconds interval for 3 different subsets and plot

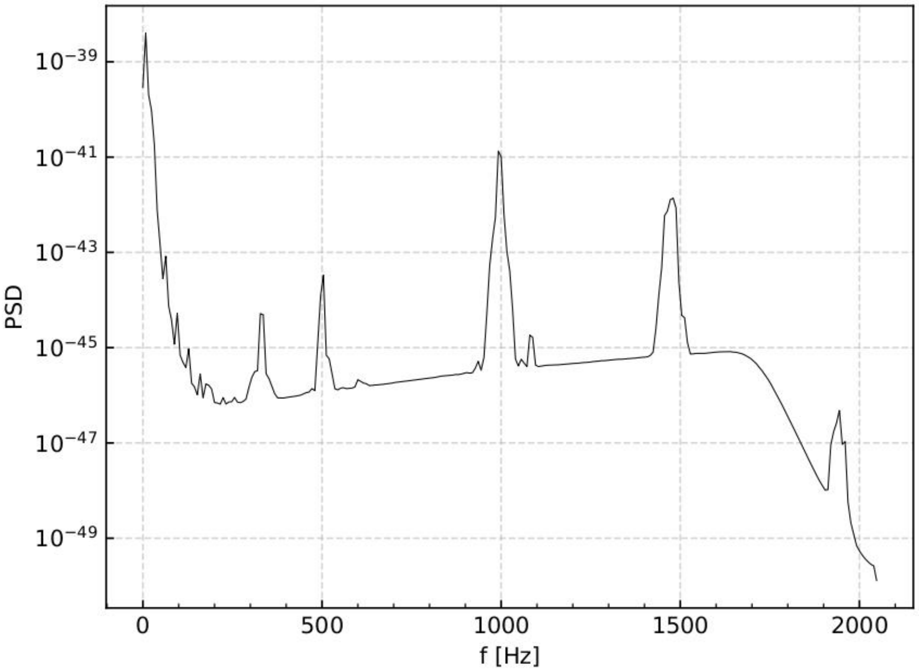
- PSDs are approximately the same, in agreement with the WSS hypothesis



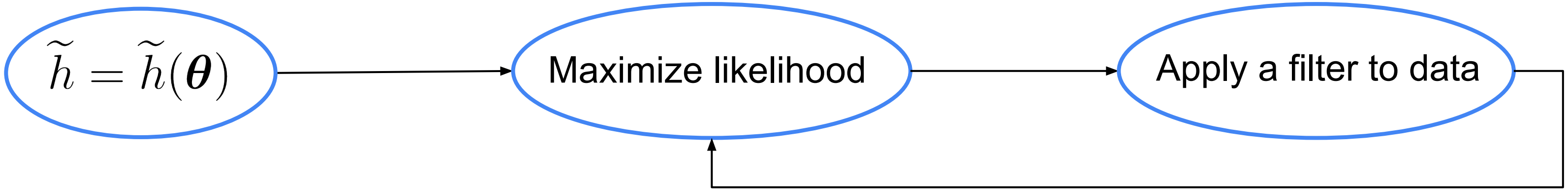
A different method

- Exploring PSD using Welch's method (`scipy.signal.welch`)
- 8 seconds chunks with 4 seconds overlap
- DFT using Tukey window and average over chunks





But how can we extract our signal?



Which filter should we apply to maximize likelihood?

Wiener Filter

We see that

$$\log \mathcal{L} = -\frac{|\tilde{d} - \tilde{h}|^2}{2S} = -\frac{1}{2}(d - h, d - h)$$

where the product $(\cdot, \cdot)$ is defined as

$$(a, b) = 4\mathrm{Re} \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} \, df$$

The mixed term (d, h) is the so called *matched* filter, or Wiener Filter.

Signal-to-noise ratio

- Matched filter is the quantity to maximize to get the most likely parameters

$$(d, h) = 4\mathrm{Re}\int_0^\infty \frac{\tilde{d}(f)\widetilde{h_{\boldsymbol{\theta}}}^*(f)}{S_n(f)}\,df$$

- It should be compared with the optimal signal-to-noise ratio, defined as

$$(\mathrm{SNR}_{\mathrm{opt}})^2 = 4\int_0^\infty \frac{|\widetilde{h_{\boldsymbol{\theta}}}(f)|^2}{S_n(f)}\,df$$

- Optimal parameters correspond to normalized SNR $\quad \rho = \dfrac{(d, h)}{\mathrm{SNR}_{\mathrm{opt}}}$

Signal model

- As suggested we use a TaylorF2 signal template to 2PN order

$$Ae^{i\psi^{(F2)}(f)}$$

- It depends on chirp mass and mass ratio

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \qquad q = \frac{m_1}{m_2}$$

$$\psi_{3.5}^{(\mathrm{F2})}(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{128\,\nu\,v^5}\left[1 + \frac{20}{9}\left(\frac{743}{336} + \frac{11}{4}\nu\right)v^2 - 16\pi v^3 + 10\left(\frac{3058673}{1016064} + \frac{5429}{1008}\,\nu + \frac{617}{144}\,\nu^2\right)v^4\right]$$

- Mass and frequency dependence in amplitude is neglected

$$\nu = \frac{q}{(1+q)^2} \qquad M = \mathcal{M}\nu^{-3/5} \qquad v = (\pi M f)^{1/3}$$

Inference

Thanks to the Bayes theorem we have posterior *pdf* on parameters

$$P(\mathcal{M}, q|d) \propto P(\mathcal{M}, q)P(d|\mathcal{M}, q)$$

and we are able to get the most probable parameters by

$$\text{argmax } P(\mathcal{M}, q|d)$$

Since we already have the likelihood function, we need prior distribution on signal parameters.

What we already know?

- Our prior knowledge about black holes masses m_1 and m_2

- We must reparameterize using chirp mass and mass ratio

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \qquad\qquad q = \frac{m_1}{m_2}$$

- We are able to change variables thanks to the formula

$$P(\boldsymbol{y}) = P(\boldsymbol{x}(\boldsymbol{y}))|\mathrm{det} J_{f^{-1}}|$$

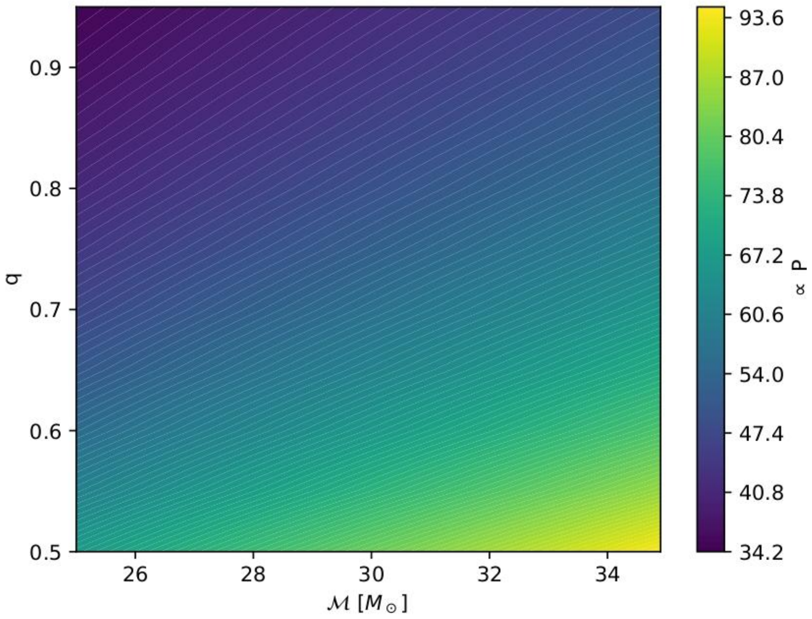
Joint prior

- Relation between new and old parameters

$$\begin{cases} m_1 = \mathcal{M}\left[q^2(1+q)\right]^{1/5} \\ m_2 = \mathcal{M}\left(\frac{1+q}{q^3}\right)^{1/5} \end{cases}$$

- Determinant of the Jacobian of the transformation

$$|\det J_{-1}| = \mathcal{M}\left(\frac{1+q}{q^3}\right)^{2/5}$$



- Assuming a uniform prior on the two masses, the joint prior is

$$\boxed{P(\mathcal{M}, q) \propto \mathcal{M}\left(\frac{1+q}{q^3}\right)^{2/5}}$$

Marginalized priors