



FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

Matching Filter for event detection

Report

UNIVERSITY OF CHILE
FACULTY OF PHYSICAL AND MATHEMATICAL SCIENCE

AUTHOR: FRANCISCO MUÑOZ
PROFESSORS: FRANCISCO FÖRSTER
PABLO HUIJSE

LAST UPDATE: MONDAY 18TH JUNE, 2018
SANTIAGO, CHILE

Abstract

Por definir.

Table of Contents

1. Introduction	1
2. Fourier Transform	1
2.1. Properties of the Fourier Transform	1
2.2. Some useful Fourier Transforms	3
2.3. Window Function	3
2.3.1. Fixing periodicity	3
2.4. Nyquist-Shannon Sampling Theorem	4
2.5. Nyquist frequency	4
2.6. Discrete Fourier Transform	5
2.7. Frequency Sampling	5
3. Fast Fourier Transform	6
3.1. Non-uniform Fast Fourier Transform	6
4. Fourier Transform as Signal Decomposition	6
4.1. What is a Signal Decomposition?	7
4.1.1. Ridge Regression	7
4.2. Fourier Dictionary	8
5. Signal Detection	8
5.1. Whitening Filter	9
5.2. Matched Filter	10
5.2.1. scopes and limitations	11
5.3. Whitened Matched Filter	11
6. Statistical Theory of Signal Detection	12
6.1. Hypothesis test	12
6.2. likelihood ratio	13
6.3. Neyman-Pearson Detection	14
6.4. NP Detection for Whitened Matched Filter	14

1. Introduction

por definir.

2. Fourier Transform

The Fourier Transform it's a Mathematical tool that allow us to decompose singals by a given dictionary or in other words, to apply a change of basis to a given signal. Thinking on a continuous space for a signal $s(t)$, it's Fourier Transform is given by the following integral, where $i = \sqrt{-1}$ denotes the imaginary unit:

$$\tilde{s}(f) = \int_{-\infty}^{\infty} s(t)e^{-2i\pi ft} dt = \mathcal{F}[s(t)] \quad (2.1)$$

And the inverse transform is:

$$s(t) = \int_{-\infty}^{\infty} \tilde{s}(f)e^{2i\pi ft} df = \mathcal{F}^{-1}[\tilde{s}(f)] \quad (2.2)$$

2.1. Properties of the Fourier Transform

Some usefull properties of the Fourier Transform are:

Fourier Transform is a linear operator. This means that, given a constant a and functions $s(t)$, $h(t)$, we can write:

$$\mathcal{F}[s(t) + h(t)] = \mathcal{F}[s(t)] + \mathcal{F}[h(t)] \quad (2.3)$$

$$\mathcal{F}[as(t)] = a\mathcal{F}[s(t)] \quad (2.4)$$

A time-shift impart a phase in the fourier Transform. For a given function $s(t)$, it's satisfied:

$$\mathcal{F}[s(t - t_0)] = \mathcal{F}[s(t)]e^{-2i\pi ft_0} \quad (2.5)$$

also can be obtained a conjugate of the transform by:

$$\mathcal{F}[s(t_0 - t)] = \mathcal{F}[s(t)]^* e^{-2i\pi ft_0} \quad (2.6)$$

The Power Spectrum equal to square of absolute value of the Fourier Transform. This computation allow us to remove the phase and the complex part which is changed to real by the absolute function.

$$\mathcal{P}_s = |\mathcal{F}[s(t)]|^2 \quad (2.7)$$

Convolution Theorem. The definition of convolution for two functions $s(t)$ and $h(t)$, is given by:

$$[s(t) * h(t)] = [h(t) * s(t)] = \int_{-\infty}^{\infty} s(t - \tau)h(\tau)d\tau \quad (2.8)$$

Calculating the Fourier Transform of this convolution leads to:

$$\mathcal{F}[s * h] = \mathcal{F}[s] \cdot \mathcal{F}[h] \quad (2.9)$$

Also, Fourier Transform of a point-wise multiplication of two functions leads to a convolution of Fourier Transforms:

$$\mathcal{F}[s(t) \cdot h(t)](f) = \mathcal{F}[s](f) * \mathcal{F}[h](f) \quad (2.10)$$

In practice using Fourier to calculate a convolution is faster than just calculate the convolution, that's why this is a very important concept.

Parseval's Formula, it says that for $s(t)$ and $h(t)$ square-integrable functions with fourier transforms given by (2.2) and (2.1), it's satisfied:

$$\int_{-\infty}^{\infty} s(t)h^*(t)dt = \int_{-\infty}^{\infty} \tilde{s}(f)\tilde{h}^*(f)df \quad (2.11)$$

Plancherel Theorem, it's a direct consequence of the Parseval Formula but in the sence of energy:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |\tilde{s}(f)|^2 df \quad (2.12)$$

Correlation. The cross-correlation of two real signals $s(t)$ and $h(t)$ can be defined in the sense of a convolution by:

$$R_{hs}(t) = h \star s = h^*(-t) * s(t) = s(t) * h^*(-t) \quad (2.13)$$

We are more interested in real signals when $s^*(-t) = s(-t)$ but the calculations will be in the general case. Using (2.8) we get:

$$R_{hs}(t) = \int_{-\infty}^{\infty} s(\tau - t)h^*(-t)d\tau \quad (2.14)$$

Applying a simple change of variables($t \longrightarrow -t$) we get:

$$R_{hs}(t) = \int_{-\infty}^{\infty} s(\tau + t)h^*(t)dt \quad (2.15)$$

Using the Time shift property (2.5) with the Parseval's Formula (2.11):

$$R_{hs}(t) = \int_{-\infty}^{\infty} \tilde{s}(f)\tilde{h}^*(f)e^{2i\pi f\tau}df \quad (2.16)$$

And this can be resumed as Fouriers Transforms:

$$R_{hs}(t) = \mathcal{F}^{-1}[\mathcal{F}[s]\mathcal{F}[h]^*] = \mathcal{F}^{-1}[\mathcal{F}[s(t) \star h^*(-t)]] \quad (2.17)$$

A particular behavior is when calculating the auto-correlation of a signal, which correspond to the Inverse Fourier Transform of the Power Spectral Density:

$$R_{hh} = \mathcal{F}^{-1}[|\tilde{h}(f)|^2] = \mathcal{F}^{-1}[\mathcal{P}_h(f)] \quad (2.18)$$

2.2. Some useful Fourier Transforms

Fourier Transform of a Sinusoid is a Delta Function. need to insert a figure in this section

$$\mathcal{F}[\cos(2\pi f_0 t)] = \frac{1}{2}[\delta(f - f_0) + \delta(f + f_0)] \quad (2.19)$$

Fourier Transform of a Dirac Comb is a Dirac Comb. The Dirac Comb is an infinite sequence of Dirac delta function placed at even intervals of size T :

$$III_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT) \quad (2.20)$$

thus, the fourier Transform correspond to a Dirac Comb in the frequency domain:

$$\mathcal{F}[III_T(t)] = \frac{1}{T} III_{1/T}(f) \quad (2.21)$$

2.3. Window Function

The Continuous Fourier Transform use the particularity that continuous functions are well defined for all times ($-\infty < t < \infty$). This is just an idealization that doesn't work on real signals because this only involve a finite span of time with some finite rate of sampling. The Fourier Transform of a discrete measured data will no longer be just the transform of the continuous underlying function, but rather the transform of the point-wise product of the signal($s(t)$) and the observing window($W(t)$), which will be a rectangular window of dirac comb.

$$s_{obs}(t) = s(t)W(t) \quad (2.22)$$

Using the convolution theorem (ec. 2.10) the transform of eq. 2.22 will be:

$$\mathcal{F}[s_{obs}(t)] = \mathcal{F}[s(t)] * \mathcal{F}[W(t)] \quad (2.23)$$

2.3.1. Fixing periodicity

Fourier Transform works only for periodic signals, this means, for signals where its first point has exactly the same value as the last point (continuity in edges). But in real world, a measured signal has always some grade of noise which lead to a discontinuity in the edges. In order to fix this discontinuity we use a window function.

For these cases, the window function will not be a rectangular, or at least, usually we don't want to use this kind of window. The most useful windows are kind-of curve like *Hann window*, *Blackmann window* or *Tukey window*. We are going to use mostly the *Tukey window* which is: **should use a example plot here for all windows**.

$$w(n) = \begin{cases} \frac{1}{2} \left[1 + \cos \left(\pi \left(\frac{2n}{\alpha(N-1)} - 1 \right) \right) \right] & 0 \leq n < \frac{\alpha(N-1)}{2} \\ 1 & \frac{\alpha(N-1)}{2} \leq n \leq (N-1)(1 - \frac{\alpha}{2}) \\ \frac{1}{2} \left[1 + \cos \left(\pi \left(\frac{2n}{\alpha(N-1)} - \frac{2}{\alpha} + 1 \right) \right) \right] & (N-1)(1 - \frac{\alpha}{2}) < n \leq (N-1) \end{cases} \quad (2.24)$$

Here, α define the form of the window, when $\alpha = 0$ we get a rectangular window and for $\alpha = 1$ we get a *Hann window*.

2.4. Nyquist-Shannon Sampling Theorem

sampling is a process of converting a signal (function-like of continuous time and/or space) into a numeric sequence, thus the sampling theorem states that, if a function $f(x)$ contains no frequencies higher than B hertz, it is completely determined by giving its ordinates at a series of points spaced $1/(2B)$ seconds apart. This means that a sufficient sample-rate for uniform sampling is therefore anything larger than $2B$ samples per second, or from other perspective, for a given sample rate f_s , **perfect reconstruction is guaranteed possible** for a bandlimit $B < f_s/2$.

In the most general case of non-uniform sampling, this theorem states that a band-limited signal can be perfectly reconstructed from its samples if the **average sampling rate** satisfies the Nyquist condition. this means that even if uniformly spaced samples can be easier reconstructed (easier algorithms), it is not a necessary condition for perfect reconstruction.

2.5. Nyquist frequency

The nyquist frequency is defined as half of the sampling rate of a discrete signal when we have evenly-sampled data

$$\nu_N = 0.5f_s \quad (2.25)$$

And correspond to the highest frequency limit of one interval over which a periodogram is uniquely defined. If we go far away from this nyquist limit, the periodogram (and the simple Fourier Transform) will have repeated information.

For unevenly-sampled data, the symmetry that allow to define this frequency as eq. 2.25 is broken and using an average sampling rate as in the Sampling Theorem doesn't work. In order to find a best approximation to the real nyquist frequency (called *pseudo-nyquist frequency*) there are several methods in literature. The one implemented here correspond to the method developed by Koen [6] who give a calculation formula to be solved for the nyquist frequency for arbitrary time spacing of measurements, this is the smallest positive root of

$$SS(\nu) = \sum_{l=1}^{N-1} \sum_{k=l+1}^N [\sin 2\pi\nu(t_k - t_l)]^2 = 0 \quad (2.26)$$

This calculation will take $O(N_\nu N^2)$ where N_ν correspond to the number of frequencies to try in the formula. Since is a very expensive calculation, generally it is more useful to know what is the maximum frequency that we are interested in.

2.6. Discrete Fourier Transform

We can pass from continuous space to discrete space by applying a Rectangular Window with a dirac comb (eq. 2.23) This is because the delta serve to collapse the Fourier Integral into a Fourier Sumatory. Considering the most general case, where a continuous signal $s(t)$ is observed at an **irregular sampling** of size N_t , leads to dirac comb:

$$III_{\{t_j\}}(t) = \sum_{j=1}^{N_t} \delta(t - t_j) \quad (2.27)$$

Using this, the direct transform will be:

$$\begin{aligned} \mathcal{F}[s_{obs}(t)] &= \tilde{s}_k = \mathcal{F}[s(t) III_{\{t_j\}}(t)] \\ \tilde{s}_k &= \int_{-\infty}^{\infty} s(t) \sum_{j=1}^{N_t} \delta(t - t_j) e^{-2\pi i f t} dt \\ \tilde{s}_k &= \sum_{j=1}^{N_t} \int_{-\infty}^{\infty} s(t) \delta(t - t_j) e^{-2\pi i f t} dt \\ \tilde{s}_k &= \sum_{j=1}^{N_t} s_j e^{-2\pi i t_j k \Delta f} \end{aligned} \quad (2.28)$$

Following the same idea we can get the inverse transform, considering that the **frequencies will always be in a regular grid** so $f_k = k \Delta f$:

$$s_j = \Delta f \sum_{k=1}^{N_f} \tilde{s}_k e^{2\pi i t_j k \Delta f} \quad (2.29)$$

the total number of frequencies (N_f) and how we define the Δf will be explained more clearly in the next sections.

2.7. Frequency Sampling

Since we are considering the most general case of Fourier Transform for non-uniform sampling, it's necessary to define clearly the frequency grid, more precisely we need to define the frequency limit on the low end and the high end, in addition to define the grid spacing.

low-end frequency limit. For a set of observations in a time interval of length T , a signal with frequency $1/T$ will complete exactly one oscillation in the observing window. Choosing a low-end frequency limit could be this frequency or just set the minimum frequency to zero for simplicity. Also is possible to give a custom low-end limit based on what frequencies we want to explore. If we have complex signals we will need to define a negative low-end frequency limit.

high-end frequency limit. In order to not miss information that could be relevant, it's necessary to look for the maximum frequency that could give us information, this is given by the nyquist frequency

defined in section 2.5. Also is posible to give a custom high-end limit based on what frequencies we want to explore.

grid spacing. Choose the correct grid spacing it's also an important point, if we use a too fine grid will leads to unnecessarily long computational times, while, use a grid too coarse leads to missing narrow peaks that fall between grid points. A razonable idea is to choose grid spacing smaller than the spected width of the periodogram peaks, these peaks has a widht of $\sim 1/T$ where T is the duration of the window function. To ensure that our grid sufficiently samples each peak, we can over-sample by some factor (n_0) using a grid spacing of

$$\Delta f = \frac{1}{n_0 T} \quad (2.30)$$

And with this we can know the total number of frequencies to evaluate

$$N_f = n_0 T (f_{max} - f_{min}) \quad (2.31)$$

choose the value of n_0 is kind-of arbitrary and we can find uses in literature in the range of $n_0 = 5$ to $n_0 = 10$ [11].

3. Fast Fourier Transform

The normal Discrete Fourier Transform showed in section 2.6 takes times $O(N^2)$ to compute wheter is the direct or the inverse transform, this is a very expensive computation and can be improved to $O(N \log(N))$ usign the algorithm of the Fast Fourier Transform(FFT). NEED TO READ MORE ABOUT THE ALGORITHM

3.1. Non-uniform Fast Fourier Transform

The same idea used in the optimization of the Discrete Fourier Tranform for evenly-sampled data called FFT can be used for uevnely-sampled data and we call this Non-unifor Fast Fourier Transform, using the same considerations. NEED TO EXPLORE MORE THE ALGORITHM

4. Fourier Transform as Signal Decomposition

An alternative approach to realize the Fourier Transfrom is to do a Signal Decomposition by regression, here we just take our data and find what coefficients give the best fit to the original data given a Dictionary which can be interpreted as a change of basis, if we use for the Dictionary the fourier Matrix, then the coefficients of the signal decomposition will be the values of the Fourier Transform.

4.1. What is a Signal Decomposition?

given a signal $s \in \mathbb{R}^N$ we want to find a representative set of coefficients $\beta \in \mathbb{R}^M$ such that:

$$\hat{s} = \Phi \hat{\beta} \quad (4.1)$$

Where $\Phi \in \mathbb{R}^{N \times M}$ is known as the dictionary matrix. Every column of this dictionary matrix correspond to an *atom*, i.e. the constituents of signal s . A dictionary is called *complete* when $M = N$. Φ can be either pre-defined or learned from data. We will select the atoms previously, for example, in a Fourier Dictionary we choose the atoms by knowing what frequencies we want.

Solve eq. 4.1 for a complete dictionary that is invertible means find $\hat{\beta}^*$ such that:

$$\hat{\beta}^* = \Phi^{-1} \hat{s} \quad (4.2)$$

When the matrix is not invertible, the solution to eq. 4.1 is obtained by solving the Least Square problem:

$$\min_{\hat{\beta}} \frac{1}{2} \|\hat{s} - \Phi \hat{\beta}\|_2^2 \quad (4.3)$$

Where $\|\cdot\|_2^2$ means the square of the L2-norm, the solution to this problem is:

$$\hat{\beta}^* = (\Phi^T \Phi)^{-1} \Phi^T \hat{s} \quad (4.4)$$

Which requires $(\Phi^T \Phi)$ to be invertible, means, having linear independent columns. The matrix $(\Phi^T \Phi)^{-1} \Phi^T$ correspond to the left Moore-Penrose pseudo-inverse of Φ .

Then complete dictionary is not often the better dictionary, here we are more interested in the overcomplete dictionary, this is a dictionary with more atoms than data samples ($M > N$). For overcomplete dictionaries eq. (4.1) is under-determined and many solutions may exists. When Φ is overcomplete, the product $(\Phi^T \Phi)$ is always a singular matrix and then (4.4) cannot be compute. In order to get an optimal its necessary to include a regularization coefficient which impose conditions to help choose one of the many solutions of the underdetermined system. There are two kind of regularization usually used, one is smooth-ness and the other is sparsity, the first refers to the least-rough or least-complex solution, this condition usually leads to dense solutions with almost all points non-zeros, and the second one refers to solution with less non-zero points and thus a more sparse solution. here we will show many method that try to find a good β .

4.1.1. Ridge Regression

this is an example of a regression that use as regularization smooth-ness, the ridge regression is defined as:

$$\min_{\beta} \alpha \|\beta\|_2^2 + \|s - \Phi \beta\|_2^2 \quad (4.5)$$

Where $\alpha > 0$ is the trade-off between reconstruction error and smooth-ness of the solution. The solution to 4.5 is:

$$\beta^* = (\Phi^T \Phi + \alpha I)^{-1} \Phi^T s \quad (4.6)$$

UNCOMPLETE, NEED TO EXPLAIN MORE METHODS

4.2. Fourier Dictionary

The discrete Fourier Transform from section 2.6 can be interpreted as a change of basis of the signals which means change the time representation of the signal \vec{s} to a frequency representation $vec\beta$ using a sine/cose basis. This will means that we can write:

$$\vec{s} = \Phi \cdot \vec{\beta} \quad (4.7)$$

Where Φ is the Dictionary/Basis to use in the change of representation. This Dictionary is the Fourier Matrix where every columnn (atom ϕ) correspond to a sine/cosine wave of a given frequency $k\Delta f$:

$$\phi_k = e^{2\pi i t_j k \Delta f} \forall t_j \quad (4.8)$$

and in order to represent Direct and Inverse Fourier Transform Correctly it's neccessary to use an orthogonal Basis which means an square matrix Dictionary with a defined Inverse, if this matrix is defined correctly (ofter in the evenly-sampled case) his adjoint will be his inverse in a way that allow us to define the Adjoint Transform and use it as the Inverse Transform.

$$\vec{\beta} = \Phi^T \cdot \vec{s} \quad (4.9)$$

Sadly for unevenly-sampled data we usually want to compute a frequency grid higher than the time grid producing a oversampling Dictionary, and even if we choose the same number of frequencies than times, there is no guarantee that the resulting Fourier Matrix will be unitary and invertible in a way that his adjoint transform be his Inverse Transform. Because of this, we need to look for a way to compute the Pseudo-inverse of the dictionary (Φ^+) instead of the Inverse, usually we can compute this by Single Value Decompostion, where the left decomposition (known as):

$$\Phi^+ \quad (4.10)$$

UNCOMPLETE

5. Signal Detection

Once we know how to interprete correctly some datas in frequency domain, using Non Unifor Fast Fourier Transform or Signal Decomposition, we are ready to work on the signal detection, i.e., on the search of a known signal in the observed data. This idea comes from the idealization of our data as an underlying signal plus additive **Wide-Sense Stationary** (WSS) Gaussian noise, then it is posible, through use of linear filters, to find what is this underlaying signal by comparing the observed data to known waveforms. Linears filters are defined as convolution of input data $x(t)$ with an impulse response filter wich is just the time reverse of the template waveform $h(t)$, for this we define the product $(x|h)$:

$$(x|h)_{t_0} = [x(t) * h(-t)](t_0) = \int_{-\infty}^{\infty} \tilde{x}(f) \tilde{h}^*(f) e^{2\pi i f t_0} df \quad (5.1)$$

Here we have used eq. (2.13) to write this a correlation. It's also important to notice that in practice we should do a *causal* time reverse for the impulse response $h(-t) \rightarrow h(T-t)$ where T is the total duration of the template waveform, but if we always use templates of same duration as the data window, then this

factor is not important due to the periodicity in Fourier Transform leadin to $h(T - t) = h(-t)$.

And the optimal Filter is just find the filter that optimize a given quantity. The theory that we are going to use correspond to the **Matched Filter** (MF) which maximize the **Signal-to-Noise Ratio** (SNR) of the observed data, and in order to do this we need to consider an input signal to the filter composed of **only White Gaussian Noise** which is a WSS Gaussian noise but with constant spectral density.

Obviously the raw data detected by any deviced will have mostly colored Gaussian Noise, this means, a WSS Gaussian noise with non-constant spectral density, then it is necessary to do a previous filter which is called *Whitening Filter* (WF) to clear the noise.

5.1. Whitening Filter

For an observed data $x(t)$ of additive WSS Gaussian colored noise $n(t)$, which could have or not an underlying signal $s(t)$, we define the whitening Filter with impulse response $h_w(-t)$ as the optimal filter when its output due to the input noise $n(t)$ is white, with variance σ^2 . This can be write as:

$$(x|h_w)_{t_0} = (s|h_w)_{t_0} + (n|h_w)_{t_0} \text{ such that } \text{var} = \mathbb{E}[(n|h_w)^2] = \sigma^2 \quad (5.2)$$

Working a little with this variance we will get:

$$\begin{aligned} \mathbb{E}[|(n|h_w)|^2] &= \mathbb{E}[(n|h_w)(n|h_w)^*] \\ &= \mathbb{E} \left[\int_{-\infty}^{\infty} n(t_0 + t_1) h_w^*(t_1) dt_1 \left(\int_{-\infty}^{\infty} n(t_0 + t_2) h_w^*(t_2) dt_2 \right)^* \right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_w^*(t_1) \mathbb{E}[n(t_0 + t_1) n(t_0 + t_2)^*] h_w(t_2) dt_1 dt_2 \end{aligned} \quad (5.3)$$

In order to work with the covariance $\mathbb{E}[n(t_0 + t_1) n(t_0 + t_2)^*]$ we need to set some considerations, first the noise used will be a 0 mean random noise, in a way that $\mathbb{E}[n(t)] = 0$, second, the noise will be invariant in time (stationary) so $n(t + t_0) = n(t)$, with this que covariance will be reduced to:

$$\mathbb{E}[n(t_0 + t_1) n(t_0 + t_2)^*] = \mathbb{E}[n(t_1) n^*(t_2)] = k_n(t_1, t_2) = k_n(\tau) \quad (5.4)$$

Where we have used $\tau = t_1 - t_2$, $k_n(\tau)$ correspond to the auto-correlation function of the noise. **Bochner's theorem** states that the covariance function of a stationary process can be represented as the fourier transform of a positive finite measure[10]:

$$k_n(\tau) = \int S_n(f) e^{2\pi i f \tau} df \quad (5.5)$$

This transform is defined as the two-sided power spectral density being a real and positive function with no zero values. Using this on eq. (5.3) we get:

$$\begin{aligned}
\mathbb{E}[|n|h_w|^2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_w^*(t_1) k_n(t_1 - t_2) h_w(t_2) dt_1 dt_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_w(t_2) \left(\int_{-\infty}^{\infty} S_n(f) e^{2\pi i f(t_1 - t_2)} df \right) h_w^*(t_1) dt_1 dt_2 \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h_w^*(t_1) e^{2\pi i f t_1} dt_1 \right) S_n(f) \left(\int_{-\infty}^{\infty} h_w(t_2) e^{-2\pi i f t_2} dt_2 \right) df \\
&= \int_{-\infty}^{\infty} \tilde{h}_w^*(f) e^{-2\pi i f T} S_n(f) \tilde{h}_w(f) e^{2\pi i f T} df \\
&= \int_{-\infty}^{\infty} \tilde{h}_w^*(f) S_n(f) \tilde{h}_w(f) df \\
&= \int_{-\infty}^{\infty} S_n(f) |\tilde{h}_w(f)|^2 df
\end{aligned} \tag{5.6}$$

With this, we define the Transfer function of the Whitening Filter impulse response as:

$$S_n(f) |\tilde{h}_w(f)|^2 = S_w(f) \text{ such that } \sigma^2 = \int_{-\infty}^{\infty} S_w(f) df \tag{5.7}$$

this directly leads to:

$$|\tilde{h}_w(f)|^2 = S_w(f) / S_n(f) \tag{5.8}$$

In a window of a finite duration with a finite number of measures N and a White Variance of σ_w^2 , then $S_w(f) = \sigma_w^2 / N = N_0$ and then:

$$|\tilde{h}_w(f)|^2 = \frac{N_0}{S_n(f)} \tag{5.9}$$

This definition of whitening will leads to a cleared signal of White noise with variance 1 and mean 0. After this, if we want to do another linear filter, we will need to normalize the impulse response by this whitening optimal filter too in order to get match the template waveform correctly.

5.2. Matched Filter

Matched filter correspond to find the optimal linear filter for maximizing the Signal To Noise Ratio (SNR) in the presence of additive white noise of a given variance σ_w and 0 mean. For this we considers some data $x(t)$ with finite-energy which can be pure noise $x(t) = n(t)$ or a signal in addition to noise $x(t) = s(t) + n(t)$, this is taken as the input to a filter with impulse response $h_{MF}(-t)$ and use the linear filter product defined in eq. (5.1)

If we do a MF using two different templates waveform, with different energy, it make hard to say what lineal filter works better because our impulse response functions are unnormalized. This is clearly seen from the noise, on the input the noise is white but on the output the noise is convolved with the template and thus is not white anymore, this introduce a change on variance:

$$\sigma_{MF} = \mathbb{E}[|x|h_{MF}|^2] = N_0 \int_{-\infty}^{\infty} |\tilde{h}_{MF}(f)|^2 df \tag{5.10}$$

Where we have used the variance of eq. (5.6) with the constant white noise PSD $S_w(f) = N_0$. Then we just need to divide the matched filter by square root of this variance in order to get a normed SNR of normal distribution $\mathcal{N}(\mu_{MF}/\sigma_{MF}, 1)$ where μ_{MF} is:

$$\mu_{MF} = \mathbb{E}[(x|h_{MF})] = \mathbb{E}[(n|h_{MF})] + \mathbb{E}[(s|h_{MF})] = 0 + (s|h_{MF})_{t_0} \quad (5.11)$$

Here we have that $\mathbb{E}[(s|h_{MF})] = (s|h_{MF})_{t_0}$ because the linear filter product itself is an expected value of the output filter. This expected value is for a given t_0 where we suppose it is produced the maximum SNR response which is written as:

$$SNR_{t_0} = \frac{(x|h_{MF})_{t_0}}{\sigma_{MF}} \quad (5.12)$$

And the optimal matched filter is when the expected underlying waveform that comes with noise on the input signals is just equal to the template waveform. from eq. (5.11) we can use $x(t)$ instead of $s(t)$ to estimate μ_{MF}

5.2.1. scopes and limitations

Here we like to define limits on the behavior of the SNR, first we consider the situation when the input data is only white noise. Using the Cauchy-Schwarz inequality on eq. (5.12) with $|\tilde{x}(f)|^2 \approx \lambda S_n(f)$, where λ depends on the normalization and method used to estimate the PSD, we get:

$$|SNR_{t_0}| = \frac{|(x|h)_{t_0}|}{|\sigma_{MF}|} = \frac{|(n|h)_{t_0}|}{|\sigma_{MF}|} \leq \frac{\sqrt{\lambda \sigma_w} \sqrt{\int_{-\infty}^{\infty} |\tilde{h}_{MF}(f)|^2 df}}{\sqrt{N_0 \int_{-\infty}^{\infty} |\tilde{h}_{MF}(f)|^2 df}} \underset{\text{Discrete case}}{=} \sqrt{\frac{\lambda N}{\sigma_w}} \quad (5.13)$$

Here we have written N_0 as the values on the discrete case $N_0 = \sigma_w^2/N$ of N samples, we are going to continue using this but we will write the equations on continuous case. Eq. (5.13) is a limit for every SNR produced when the input data is only white noise.

Another important limit is to find the maximum value that a SNR can achieve for any type of input data (with or without underlying signal), again, using the Cauchy-Schwarz inequality, we get in this case:

$$|SNR_{t_0}| \leq \frac{\sqrt{\int_{-\infty}^{\infty} |\tilde{x}(f)|^2 df} \sqrt{\int_{-\infty}^{\infty} |\tilde{h}_{MF}(f)|^2 df}}{\sqrt{N_0 \int_{-\infty}^{\infty} |\tilde{h}_{MF}(f)|^2 df}} \underset{\text{Discrete case}}{=} \sqrt{\frac{E_x N}{\sigma_w^2}} \quad (5.14)$$

Where $E_x = \int_{-\infty}^{\infty} |\tilde{h}_{MF}(f)|^2 df = \int_{-\infty}^{\infty} |h_{MF}(t)|^2 dt$ is the energy of the input data, here we see that the maximum value possible of the SNR only depend on the input data (energy, number of samples and white noise variance) and not on the template waveform.

5.3. Whitened Matched Filter

We can (and must) use the Whitening Filter (WF) together with the Matched Filter (MF) for our (guessed) input signal with colored WSS noise. Here we just need to apply WF to both the input data and the impulse response on the MF:

$$(x|h_{MF})_{t_0}^{WMF} = [\{x * h_w\} * \{h_{MF} * h_w\}] = [x_w * h_{WMF}] \quad (5.15)$$

Using the definition of lineal filter on eq. (5.1) with eq. (2.9), this will be:

$$(x|h_{\text{MF}})_{t_0}^{\text{WMF}} = \int_{-\infty}^{\infty} \tilde{x}(f) \tilde{h}_{\text{MF}}^*(f) |\tilde{h}_w(f)|^2 e^{2\pi i f t_0} df \quad (5.16)$$

Where $|\tilde{h}_w(f)|^2$ is given by eq. (5.9) this leads to:

$$(x|h_{\text{MF}})_{t_0}^{\text{WMF}} = N_0 \int_{-\infty}^{\infty} \frac{\tilde{x}(f) \tilde{h}^*(f)}{S_n(f)} e^{2\pi i f t_0} df \quad (5.17)$$

This also affect the variance and the expected (value) remains with same definition as shown in eq. (5.11) but for the product defined in eq. (5.17):

$$\sigma_{\text{WMF}}^2 = \mathbb{E}[|(x|h_{\text{MF}})_{t_0}^{\text{WMF}}|^2] = N_0 \int_{-\infty}^{\infty} |\tilde{h}_{\text{MF}}(f)|^2 |\tilde{h}_w(f)|^2 df = N_0^2 \int_{-\infty}^{\infty} \frac{|\tilde{h}_{\text{MF}}(f)|^2}{S_n(f)} df \quad (5.18)$$

$$\mu_{\text{WMF}} = \mathbb{E}[(x|h_{\text{MF}})_{t_0}^{\text{WMF}}] = \mathbb{E}[(x - n|h_{\text{MF}})_{t_0}^{\text{WMF}}] \quad (5.19)$$

Here (and in the next sections) for simplicity we will write $(x|h_{\text{MF}})_{t_0}^{\text{WMF}} = (x|h)$, $\sigma_{\text{WMF}} = \sigma$ and $\mu_{\text{WMF}} = \mu$. Thus the SNR for the Whitenen Matched Filter (written as ρ for simplicity) will be:

$$\rho_{t_0} = \frac{(x|h)}{\sigma} \quad (5.20)$$

This division doesn't depend on N_0 so clearly this matched filter doesn't work with White noise (and in next we will ignore this factor), instead, consider any WSS Gaussian noise. For this definition, the limit values of the matched filter defined in eq. (5.13) and (5.14) should be the same but due to numerical computations, could be not exactly the same. This SNR has the same distribution than before $\mathcal{N}(\mu/\sigma, 1)$.

6. Statistical Theory of Signal Detection

The observed signal carried an additive noise introduced by the detector and is usually modeled as a Wide-Sense Stationary (WSS) Gaussian random process, this means, the problem of extracting the signal from the noise (as shown in section (5)) can be interpreted as a statistical one. The presence of signal changes the statistical characteristics of the data observed $x(t)$, affecting his Probability Density Function (PDF).

All the theory that will be described here are explained more extense in [5, 7, 4].

We are going to work with statistical hypothesis testing and then define a likelihood ratio with a threshold value in order to decide if there is really a signal or not.

6.1. Hypothesis test

The main idea here is to consider two scenarios (binary hypothesis) one is the *null hypothesis* H_0 which consider data absent of signal and *alternative hypothesis* H_1 which consider a signal present.

$$H_0 : x(t) = n(t) \quad (6.1)$$

$$H_1 : x(t) = n(t) + s(t) \quad (6.2)$$

The *Binary Hypothesis test* or *binary decision rule* consider 4 scenarios which can be representes as a confusion matrix. Two of these fourth possibilities are consider errors:

- $P(H_0|H_0)$, this is the probability of choose H_0 when H_0 is true. It's a **correct decision**.
- $P(H_1|H_0)$, probability of choose H_1 when H_0 is true. This is a Type I error and it's called **false alarm probability**.
- $P(H_0|H_1)$, probability of choose H_0 when H_1 is true. Type II error known as **False dismissal probability**.
- $P(H_1|H_1)$, probability of choose H_1 when H_1 is true. It's a **correct decision**.

The type II error is also the *probability of detection* of the signal. For the Test scenario, the type I error is called *significance of the test* and type II error is *power of the test*.

Next, we need to find a test that is optimal. The main idea here, no matter what approach we use, is to calculate the **likelihood ratio**.

6.2. likelihood ratio

From the binary hypothesis test, for a given measure x , we will have a probability $P(H_0|x)$ of H_0 being true and probability $P(H_1|x)$ of H_1 being true, obviously, if H_1 is actually true, then $P(H_1|x) > P(H_0|x)$, same for case of H_0 being true, this can be written in the compact form:

$$P(H_1|x) \underset{H_0}{\overset{H_1}{>}} P(H_0|x) \quad (6.3)$$

Here is implicit that when $P(H_0|x) = P(H_1|x)$ we cannot know what hypothesis is correct. These conditional probabilities are using the posterior probability of the hypothesis, this is different from the *a priori* probabilities of every hypothesis $P(H_0)$ and $P(H_1)$. In order to evaluate the *posteriori* probabilities in (6.3) we use the Bayes' rule:

$$\frac{p_1 f(x|H_1)}{f(x)} \underset{H_0}{\overset{H_1}{>}} \frac{p_0 f(x|H_0)}{f(x)} \quad (6.4)$$

Where $f(x|H_j)$ is just the conditional PDF and $f(x)$ is the PDF of measure the observed data, since is the same for both hypothesis, we will not consider his values. On eq. (6.4) we have considered that $f(x|H_j) > 0$ for $j = 0, 1$. Next we can just rewrite this equation in order to get the likelihood ratio $\Lambda(x)$:

$$\Lambda(x) = \frac{f(x|H_1)}{f(x|H_0)} \underset{H_0}{\overset{H_1}{>}} \frac{p_0}{p_1} = \eta \quad (6.5)$$

Where η is a **threshold** that can be defined in several ways, one can estimate the optimal case using bayesian inference or Neyman-Pearson (NP) detection between many others. For now we will use the NP theory to define the detection and threshold.

6.3. Neyman-Pearson Detection

This approach consider a context in which we want to maximize the probability of detection defined as:

$$P_D = P(H_1|H_1) = \int_{D_1} f(x|H_1)dx \quad (6.6)$$

While keeping the probability of false alarm:

$$P_{FA} = P(H_1|H_0) = \int_{D_1} f(x|H_0)dx = \alpha \quad (6.7)$$

Below a pre-specified level α . The values in D_1 are all values for which the likelihood function is higher than a given threshold η . Then the threshold is picked to provide the largest possible P_D while ensuring that P_{FA} is not larger than the pre-specified level. The smaller the η , the larger the decision region D_1 and the value of P_D , but P_{FA} grows as well. We can just pick the smaller threshold that satisfied the given bound of P_{FA} .

6.4. NP Detection for Whitened Matched Filter

Remembering that we consider the noise as a WSS White Gaussian random process, we can modelate this noise as a normal distribution of 0 mean ($\mu = 0$) and a given variance σ ($\mathcal{N}(0, \sigma^2)$), thus, for both hypothesis from eq. (6.1) and (6.2) $n(t)$ distribute as:

$$n(t) \sim \mathcal{N}(0, \sigma^2) \longrightarrow f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (6.8)$$

When the matched filter as defined in sec. (5.3) is applied, our hypothesis tests are:

$$H_0 : (x|h) = (n|h) \sim \mathcal{N}(0, \sigma) \quad (6.9)$$

$$H_1 : (x|h) = (n|h) + (s|h) \sim \mathcal{N}(\mu, \sigma) \quad (6.10)$$

then, for the SNR we normalize by square root of the variance, leading to hypothesis tests:

$$H_0 : \frac{(x|h)}{\sigma} = \frac{(n|h)}{\sigma} \sim \mathcal{N}(0, 1) \quad (6.11)$$

$$H_1 : \frac{(x|h)}{\sigma} = \frac{(n|h)}{\sigma} + \frac{(s|h)}{\sigma} \sim \mathcal{N}\left(\frac{\mu}{\sigma}, 1\right) \quad (6.12)$$

Here the product $(s|h)$ is optimal when the signal $s(t)$ is equal to the impulse response shifted by the optimal time of match $s(t) = h(t_0 - t)$, using the base definition of linear filter (5.1) written as a simple convolution (2.8), with the parseval's equation (2.11) and the equations defined in sec. (5.3):

$$(s|h) = [h(t_0 - t) * h(-t)](t_0) = \int_{-\infty}^{\infty} h(t_0 - (t_0 - t))h(-t)dt = \int_{-\infty}^{\infty} |\tilde{h}(f)|^2 df = (h|h)_{t_0=0} = \sigma^2 \quad (6.13)$$

The expected value of the optimal filter is just the variance of the output filter. And the Probability Density Function (PDF) of normal distributions is:

$$f(\rho) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\rho-\mu)^2}{2\sigma^2}} \quad (6.14)$$

Using this PDF as a conditional PDF with the hypothesis from eq. (6.9) and (6.9), where $\mu = (s|h)$ for H_1 , the likelihood ratio (6.5) is written as:

$$\begin{aligned} \Lambda\{(x|h)\} &= \frac{(\sqrt{2\pi})^{-1} \exp\left(-\frac{((x|h)-(s|h))^2}{2\sigma^2}\right)}{(\sqrt{2\pi})^{-1} \exp\left(-\frac{(x|h)^2}{2\sigma^2}\right)} \\ &= \exp\left(\frac{2(x|h)(s|h) - (s|h)^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{(s|h)}{\mathbb{E}[(n|h)^2]} \left\{ (x|h) - \frac{1}{2}(s|h) \right\}\right) \end{aligned} \quad (6.15)$$

The factor $(s|h)/\mathbb{E}[(n|h)^2]$ on the optimal case is simplified resulting in:

$$\Lambda\{(x|h)\} = \exp\left(\left\{ (x|h) - \frac{1}{2}(h|h) \right\}\right) \quad (6.16)$$

This likelihood ratio is the same that the one used in [5, 4] even if we derived the formulas from different perspectives. If we simplify this likelihood ratio by doing a natural logarithm, we get:

$$\ln[\Lambda\{(x|h)\}] = \left\{ (x|h) - \frac{1}{2}(h|h) \right\} \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \ln \left[\frac{p_0}{p_1} \right] = \ln[\eta] \quad (6.17)$$

And here we can write a critical criteria over Threshold to decide what hypothesis is correct:

$$\begin{matrix} H_1 \\ (x|h) > \\ < \\ H_0 \end{matrix} \ln[\eta] + \frac{1}{2}(h|h) = \eta' \quad (6.18)$$

Here we see that we can write the optimal threshold if we know what is the optimal and the *a priori* probability of H_0 and H_1 (or at least is ratio), but this is not know for the mosts cases, so we need to estimate a threshold for a given constrains. Under the NP criteria, we choose this threshold η' to maximize P_D and satisfy the desired false alarm P_{FA} .

We know that the solution for eq. (6.7) is a *complementary error function*:

$$P_{FA} = \int_{\eta'}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)} dx = \text{erfc}\left(\frac{\eta'}{\sigma}\right) \quad (6.19)$$

Where $\text{erfc}()$ means the complementary error function. Then we can solve for η' which leads to:

$$\eta' = \text{erfc}^{-1}(P_{FA})\sigma \quad (6.20)$$

With this threshold defined, we calculate the probability of detection (6.6) which solution is:

$$P_D = \int_{\eta'}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} dx = \text{erfc}\left(\frac{\eta' - \mu}{\sigma}\right) \quad (6.21)$$

In terms of the False Alarm:

$$P_D = \text{erfc}\left(\frac{\sigma \text{erfc}^{-1}(P_{FA}) - \mu}{\sigma}\right) = \text{erfc}[\text{erfc}^{-1}(P_{FA}) - \mu/\sigma] = \text{erfc}\left[\text{erfc}^{-1}(P_{FA}) - \frac{(x|h)_{t_0}}{\sqrt{(h|h)_{t_0=0}}}\right] \quad (6.22)$$

This is the final value that we are going to use to determine if our SNR represent a detection or not, the parameters needed for this are the False alarm value P_{FA} and the values calculated for the filter used to estimate the mean and variance. We can also define the counterpart, the probability of False alarm for a given Probability of detection:

$$P_{FA} = \text{erfc}\left[\text{erfc}^{-1}(P_D) + \frac{(x|h)_{t_0}}{\sqrt{(h|h)_{t_0=0}}}\right] \quad (6.23)$$

What of these two probabilities we want to compute depend on what kind of event we observe.

References

- [1] B. P. Abbott et al. “GW150914: First results from the search for binary black hole coalescence with Advanced LIGO”. In: *Phys. Rev. D* 93.12, 122003 (June 2016), p. 122003. DOI: [10.1103/PhysRevD.93.122003](https://doi.org/10.1103/PhysRevD.93.122003). arXiv: [1602.03839](https://arxiv.org/abs/1602.03839) [gr-qc].
- [2] B. Allen et al. “FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries”. In: *Phys. Rev. D* 85.12, 122006 (June 2012), p. 122006. DOI: [10.1103/PhysRevD.85.122006](https://doi.org/10.1103/PhysRevD.85.122006). arXiv: [gr-qc/0509116](https://arxiv.org/abs/gr-qc/0509116) [gr-qc].
- [3] C. Cutler and É. E. Flanagan. “Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral waveform?” In: *Phys. Rev. D* 49 (Mar. 1994), pp. 2658–2697. DOI: [10.1103/PhysRevD.49.2658](https://doi.org/10.1103/PhysRevD.49.2658). arXiv: [gr-qc/9402014](https://arxiv.org/abs/gr-qc/9402014) [gr-qc].
- [4] L. S. Finn. “Detection, measurement, and gravitational radiation”. In: *Phys. Rev. D* 46 (Dec. 1992), pp. 5236–5249. DOI: [10.1103/PhysRevD.46.5236](https://doi.org/10.1103/PhysRevD.46.5236). arXiv: [gr-qc/9209010](https://arxiv.org/abs/gr-qc/9209010) [gr-qc].
- [5] P. Jaranowski and A. Królak. “Gravitational-Wave Data Analysis. Formalism and Sample Applications: The Gaussian Case”. In: *ArXiv e-prints* (Nov. 2007). arXiv: [0711.1115](https://arxiv.org/abs/0711.1115) [gr-qc].
- [6] Chris Koen. “The Nyquist frequency for irregularly spaced time-series: a calculation formula”. In: *Monthly Notices of the Royal Astronomical Society* 371.3 (2006), pp. 1390–1394. DOI: [10.1111/j.1365-2966.2006.10762.x](https://doi.org/10.1111/j.1365-2966.2006.10762.x). URL: <http://dx.doi.org/10.1111/j.1365-2966.2006.10762.x>.
- [7] A. Oppenheim and G. Verghese. *Course Notes: Signals, Systems and Inference*. Spring 2010. URL: <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-011-introduction-to-communication-control-and-signal-processing-spring-2010/readings/>.
- [8] B. J. Owen and B. S. Sathyaprakash. “Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement”. In: *Phys. Rev. D* 60.2, 022002 (July 1999), p. 022002. DOI: [10.1103/PhysRevD.60.022002](https://doi.org/10.1103/PhysRevD.60.022002). arXiv: [gr-qc/9808076](https://arxiv.org/abs/gr-qc/9808076) [gr-qc].
- [9] A. Raman. “On the Signal Processing Operations in LIGO signals”. In: *ArXiv e-prints* (Nov. 2017). arXiv: [1711.07421](https://arxiv.org/abs/1711.07421).
- [10] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006. URL: <http://www.gaussianprocess.org/gpml/>.
- [11] J. T. VanderPlas. “Understanding the Lomb-Scargle Periodogram”. In: *ApJS* 236, 16 (May 2018), p. 16. DOI: [10.3847/1538-4365/aab766](https://doi.org/10.3847/1538-4365/aab766). arXiv: [1703.09824](https://arxiv.org/abs/1703.09824) [astro-ph.IM].