Assignment 2

Alma Sehic s274208 – Group 6

Applied Signal Processing Laboratory 2023

Politecnico di Torino

Introduction

- In this assignment an exercise which makes use of correlation of two signals will be covered.

- The correlation gives a measure of the similarity between two signals:

$$C_{12} = \int s_1(t)s_2(t)\,\mathrm{d}t$$

- We know that if $s_1(t)$ and $s_2(t)$ have unitary energy their correlation coefficient will be:

$$0 \leq |c_{12}| \leq 1$$

where $c_{12} = 1$ when two signals are equal and $c_{12} = 0$ when they are orthogonal

Exercise 2.b – Detection of a pulse: energy detection vs. correlation

- A time window is observed which may or may not contain a noisy version of our generated signal
- A distinction is made between case H_0 (window contains only noise) and case H_1 (window contains a noisy version of the pulse)
- The correlation between the observed window and our pulse is computed
- If the correlation is very high, we conclude that the window contains the pulse, otherwise only noise

Exercise 2.b – Variable definition

Variables used are the following:

Sampling frequency: $f_s = 100\ Hz$

Sampling time: $T_s = \dfrac{1}{f_s} = 0.01\ s$

Window time: $T = 1\ s$

Frequency: $f_1 = 1\ Hz$

Amplitude: $A_1 = \sqrt{2}$

Time axis: $t_a = 0{:}\,T_s{:}\,T - T_s$

Exercise 2.b – Generating signal s(t)

- After generating our sinusoidal signal:

$$s(t) = A_1 \sin(2\pi f_1 t_a)$$

we can verify that it is an already normalized signal since its energy and power are equal to 1 using the following MATLAB formulae:

```
E_s = sum(s.^2)*T_s;
P_s = (A1^2)/2;
```

where the power of the signal is calculated using the theoretical formula for the power of real periodic signals

It is necessary to ensure that the signal we wish to perform a correlation with has unitary energy so the correlation coefficient will be between 1 and 0. In this case, our signal $s(t)$ satisfies this condition

Exercise 2.b – Generating noise signal $\mathrm{n}(t)$

- The SNR (signal-to-noise ratio) is set to $-10\ \mathrm{dB}$ and converted into the natural value using the formula:

$$SNR_{nat} = 10^{\frac{SNR_{dB}}{10}} = 0.10$$

- The corresponding noise power is calculated as following:

$$P_n = \frac{P_s}{SNR_{nat}}$$

- Lastly, a realization of a noise signal is generated as:

$$n(t) = \sqrt{var} * randn(1, N)$$

$$\longrightarrow \quad var = P_n$$

where $randn(1, N)$ is a MATLAB function which returns a 1 by N array of normally distributed random numbers

and where var is the variance of the noise signal. Since the noise signal is an ergodic signal, we can instead use the power P_n calculated so in this case:

$$var = P_n$$

Exercise 2.b – Plotting noisy signal r(t)

- To get our noisy signal, we take the sum of our signal and the noise signal:

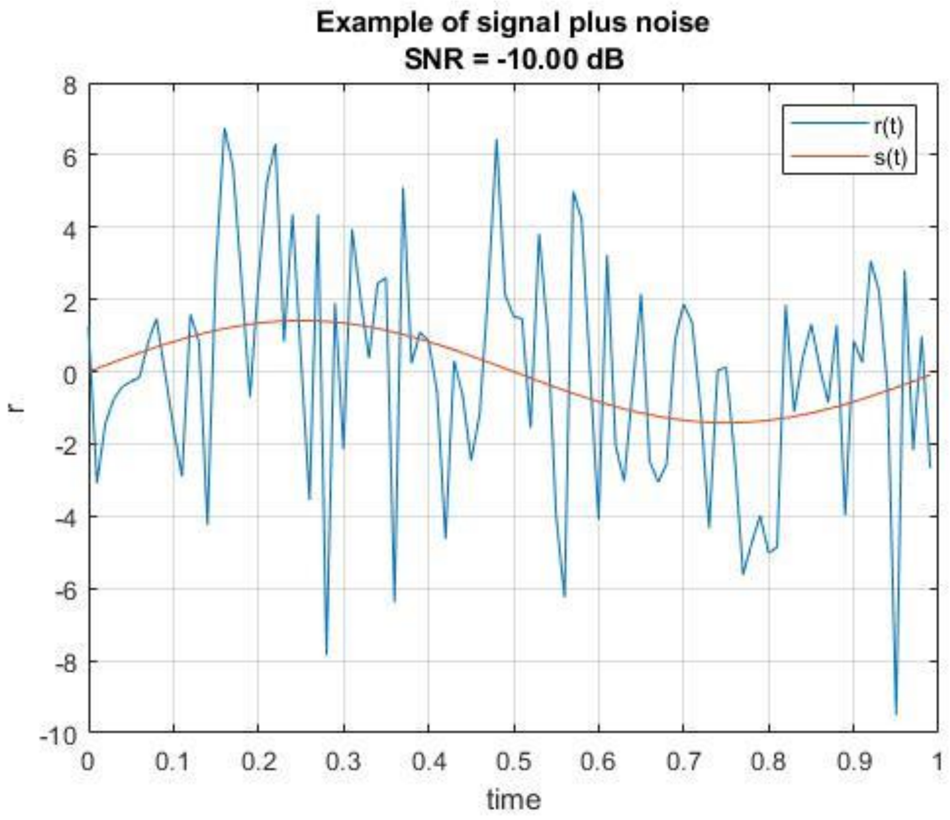

Figure 1.1: Plot of the signal plus noise signal $r(t)$ and our original signal $s(t)$

$$r(t) = s(t) + n(t)$$

After plotting our original signal $s(t)$ and the signal plus noise signal $r(t)$, it can be seen that the $r(t)$ signal is following the shape of the $s(t)$ signal. But because of the noise, the shape of the original signal with respect to the noisy signal is very hard to be identified.

Exercise 2.b – Correlation test

- A $num_{sim} = 100{,}000$ noise realizations $\mathrm{n}(t)$ is generated
- For each realization we observe $\mathrm{n}(t)$ (observed signal under H_0) and $r(t) = s(t) + n(t)$ (observed signal under H_1)
- A correlation test Γ is performed as such:

$$\boxed{\begin{aligned} \Gamma | H_0 &= |(n', s)| \\ \Gamma | H_1 &= |(r', s)| \end{aligned}}$$

where n' and r' have been normalized to have unitary energy (for each realization on the noise and noise+signal) using the expression:

$$\boxed{n'(t) = \frac{n(t)}{\sqrt{E_n}}} \qquad \boxed{r'(t) = \frac{r(t)}{\sqrt{E_r}}}$$

Exercise 2.b – Correlation test under H_0 and H_1

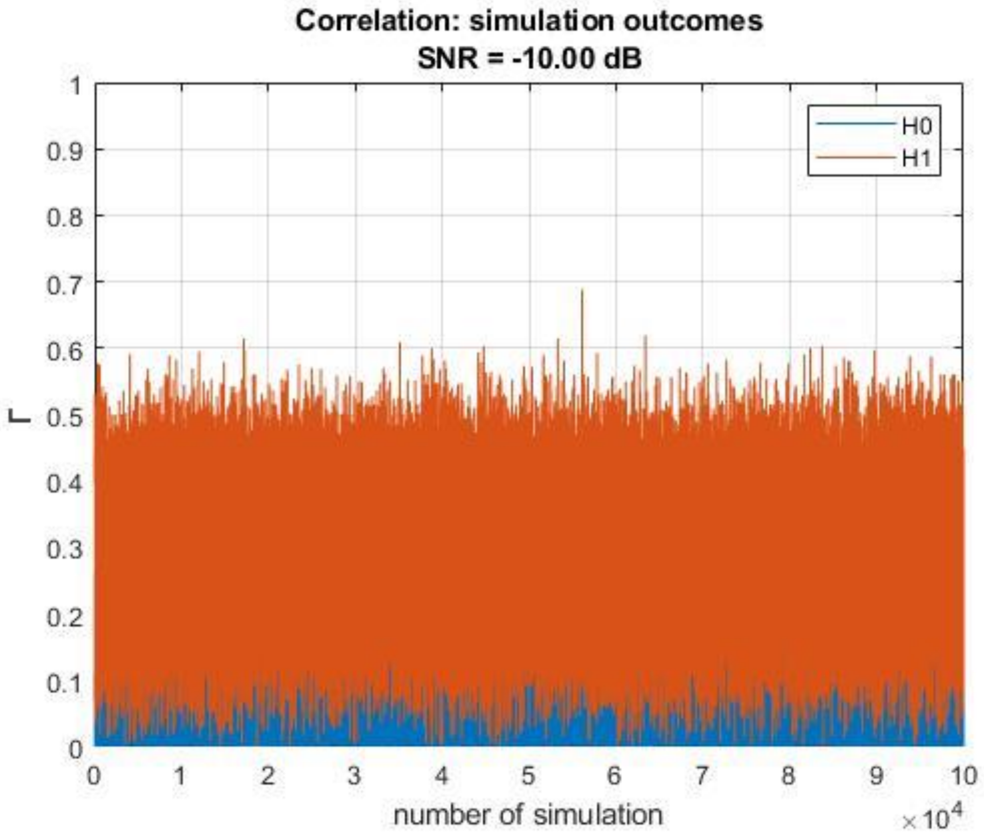

Figure 1.2: Plot of the correlation: simulation outcome Γ of the noise signal under H_0 and signal plus noise signal under H_1 wrt. the number of simulations

In Figure 1.2 it can be seen that the correlation coefficient under H_1 (original signal + noise) is way higher than the correlation test under H_0 (only noise).

This outcome was expected because we know that if the observed window in the H_1 case contains the pulse $s(t)$ the correlation coefficient will be very high, otherwise it contains only noise as in case H_0 where the correlation coefficient will be very low.

Exercise 2.b – Pdfs of Γ under H_0 and H_1

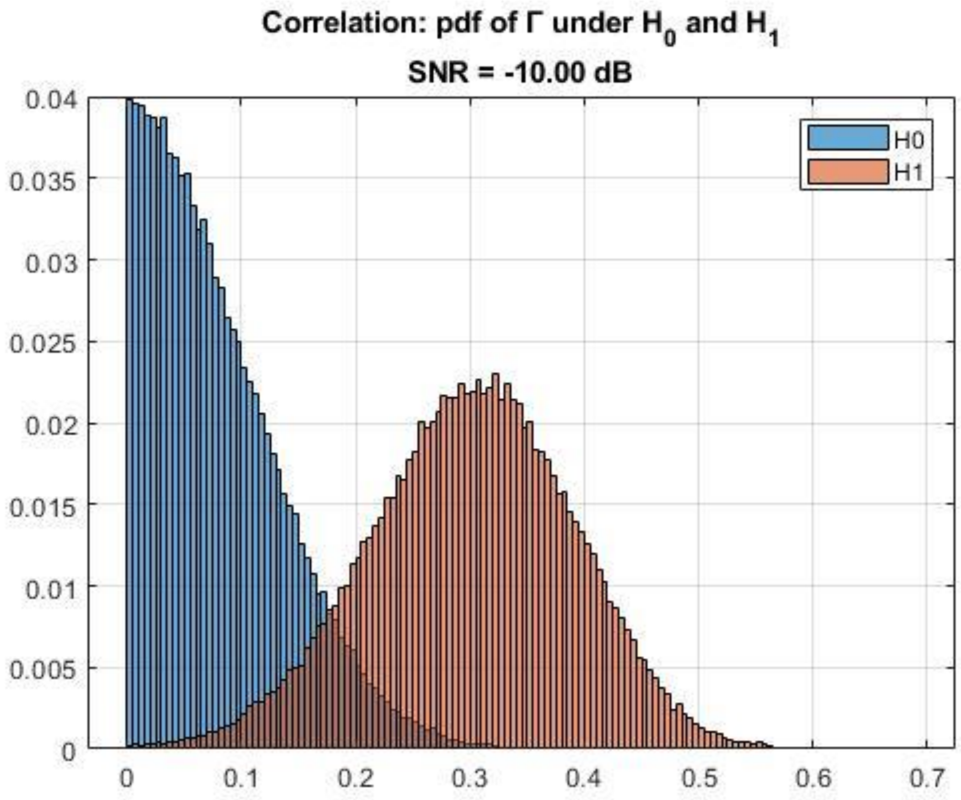

Figure 1.3: Pdfs of the correlation coefficient under H_0 and H_1

By plotting the PDFs (Probability Density Functions) of the correlation coefficient Γ under H_0 and H_1 we can observe that H_0 (due to noise only, being a random variable) has a normal distribution, therefore has the mean value equal to zero as seen in Figure 1.3, in blue.

Consequently, the correlation coefficient Γ of the noisy signal $r(t)$ (under H_1) follows the same behavior. Once again, it has a normal distribution due to the noise component, but its mean is not zero – it is shifted to the right, to approximately 0.3.

Exercise 2.b – Threshold values for computing P_{md} and P_{fa}

- It is necessary to define a threshold with respect to which we can decide if the pulse is present (is case Γ ≥ t), otherwise the pulse is absent.
- 100 threshold values between the minimum and the maximum test values of Γ are then considered and a linearly spaced vector is generated as such:

$$t_{min} = \min(H0)$$
$$t_{max} = \max(H1)$$
$$N_{bins} = 100$$
$$t2 = linspace(t_{max}, t_{min}, N_{bins})$$

where $\mathrm{linspace(x, y, N)}$ is a MATLAB function which generates N points. The spacing between the points is $\mathrm{(y - x)/(N - 1)}$

- Given the 100 threshold values, we compute the probability of a missed detection P_{md} and the probability of a false alarm P_{fa}

where *missed detection* indicates that the pulse is present but due to noise the test is below the threshold and *false alarm* indicates that the pulse is absent but due to noise the test is above the threshold

Exercise 2.b – Plot of P_{fa} and P_{md} vs the threshold values t

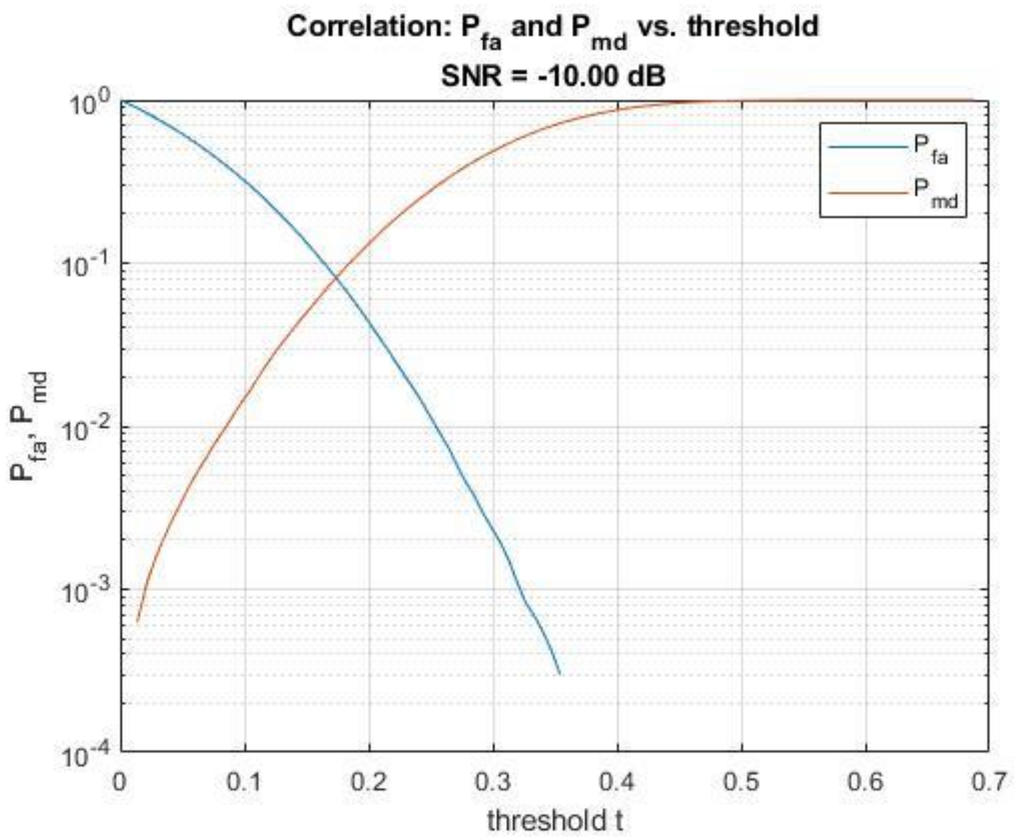

Figure 1.4: Plot of P_{fa} and P_{md} vs the threshold values t

The following probability values are calculated such that:

$$P_{fa} = \Pr(\Gamma \geq \mathrm{t} \mid H_0) \qquad P_{md} = \Pr(\Gamma < \mathrm{t} \mid H_1)$$

Since we would want these probabilities to be low, it is necessary to find a trade-off for the threshold t : if the threshold t is high, P_{fa} is low but P_{md} is high. On the other hand, if the threshold is low, P_{fa} is high but P_{md} is low.

This exact observation can be seen in Figure 1.4 where we could say that a good choice of threshold value would be at the point of the two lines crossing

Exercise 2.b – Plot of P_{fa} and P_{md} vs the threshold values t

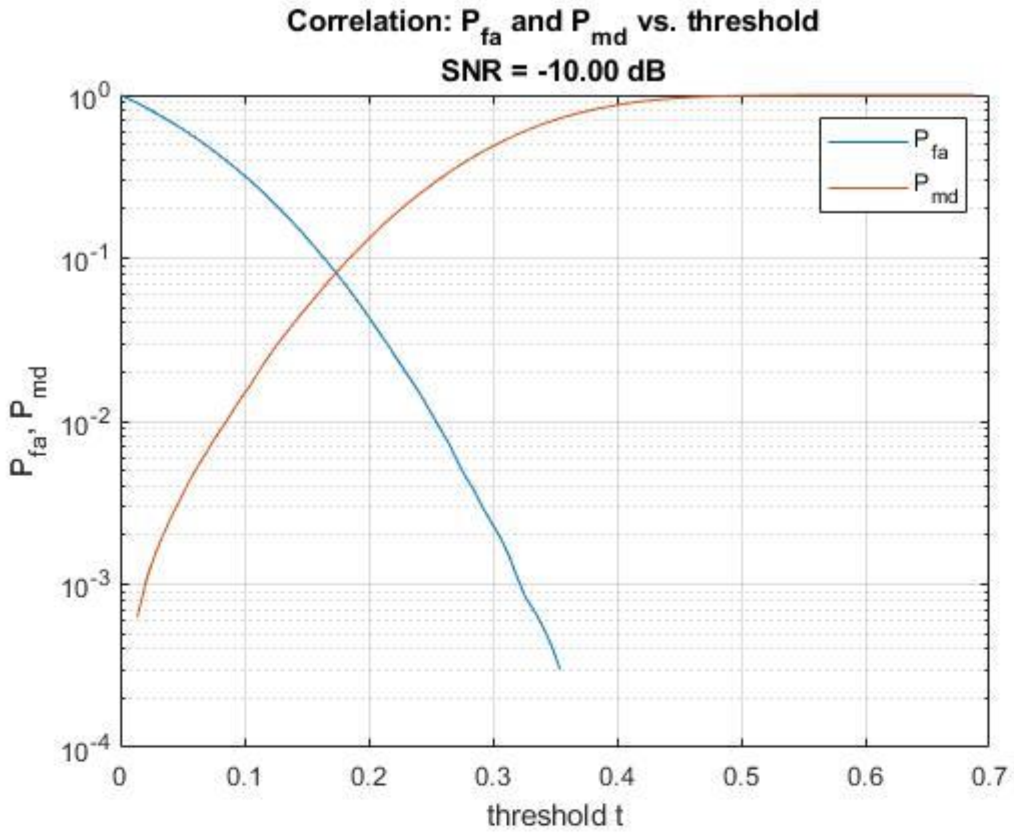

Figure 1.4: Plot of P_{fa} and P_{md} vs the threshold values t

The probabilities have been calculated using the following MATLAB example functions:

P_fa (t) = length (find (H0 >= t))/num_sim

P_md (t) = length (find (H1 < t))/num_sim

Furthermore, to have a good reliability of the measured probability, if:

length (find (H0 >= t))/num_sim < 30 set $P_{fa} = 0$

length (find (H1 < t))/num_sim < 30 set $P_{md} = 0$

Exercise 2.b – Plot of the ROC curve P_d vs P_{fa}

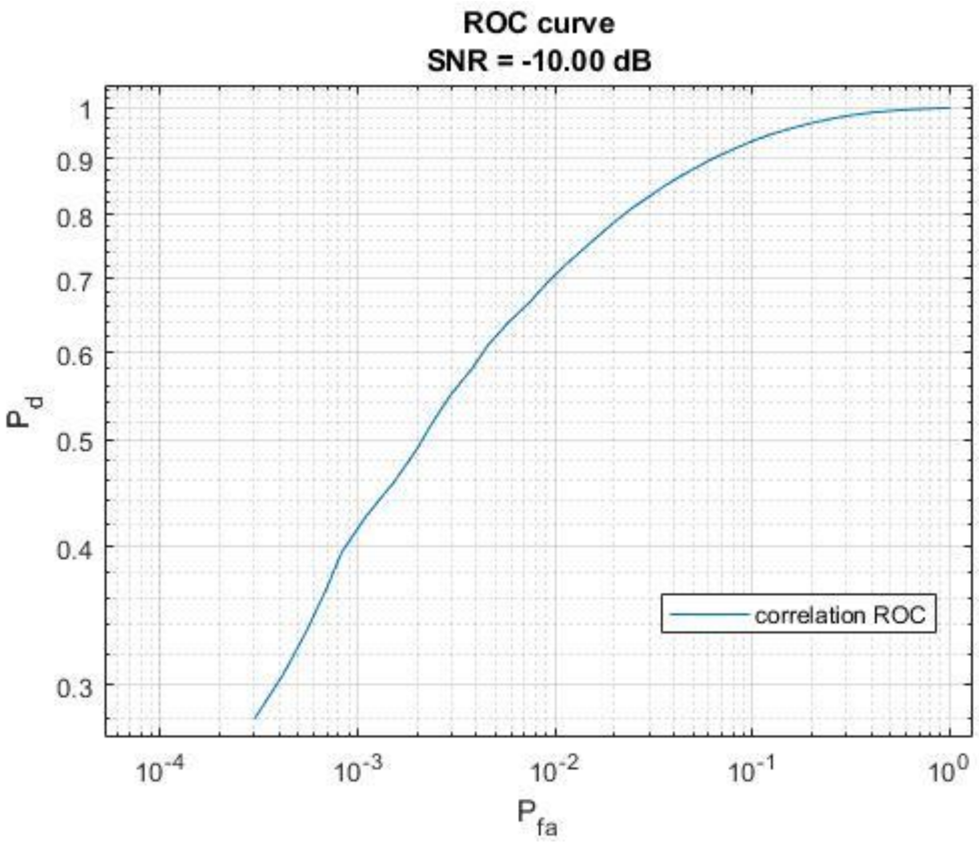

Figure 1.5: Plot of the ROC curve $P_d = (1 - P_{md})$ vs P_{fa}

- Another approach is to plot and observe the behavior of a ROC (Receiver Operating Characteristic) curve at thresholds t .
- A ROC curve is a plot of the probability of detection P_d vs probability of false alarm P_{fa} for a given SNR, where:

$$\boxed{P_d = (1 - P_{md})}$$

- The x and y axes on the ROC curve are also known as the FPR (False Positive Rate) and TPR (True Positive Rate) respectively.

where "*FPR* shows how many incorrect positive results occur among all negative samples and *TPR* shows how many correct positive results occur among all positive samples"

Exercise 2.b – ROC curve cont'd

- The two axis (FPR and TPR) represent relative trade-offs between the benefits and the costs.

- The best possible ROC curve is the one which approaches the top left corner where P_d approaches 1 and P_{fa} approaches 0 on their respective axes; whereas it has an unsatisfactory yield when the ROC curve approaches a diagonal.

- A visual representation of the effect of different SNR values on the ROC curve can be seen in the next slide

ROC curve with different SNR values - example 1: larger SNR

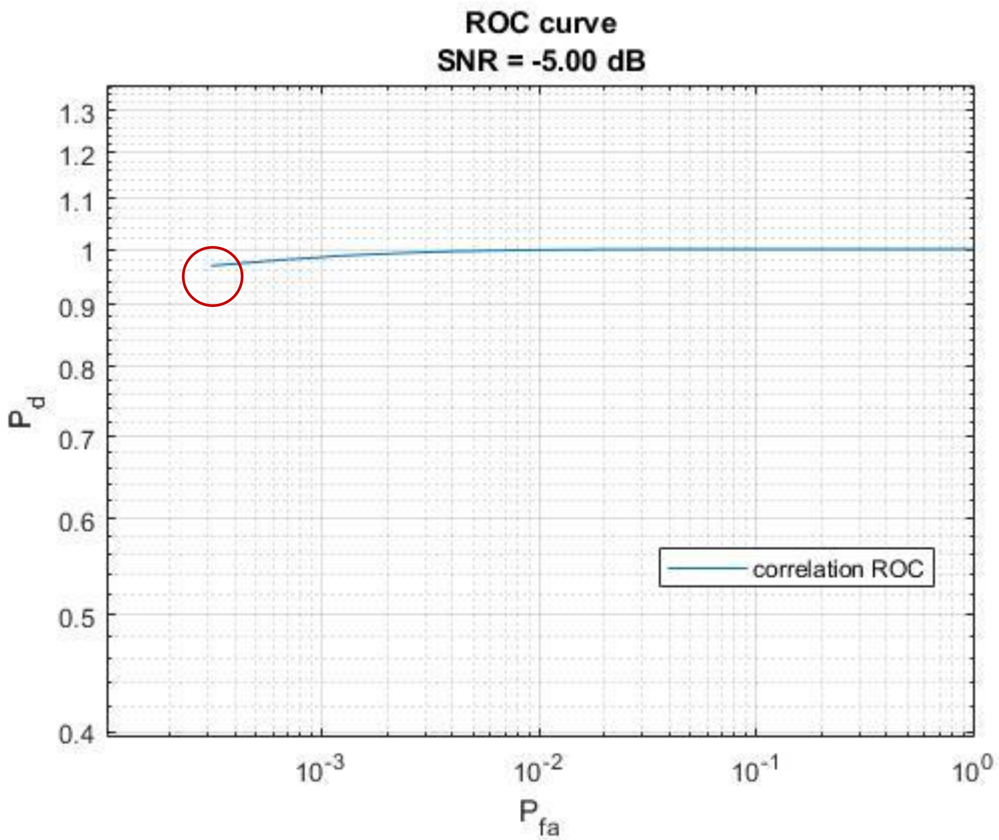

Figure 1.6.1: Plot of the ROC curve with SNR = -5 dB

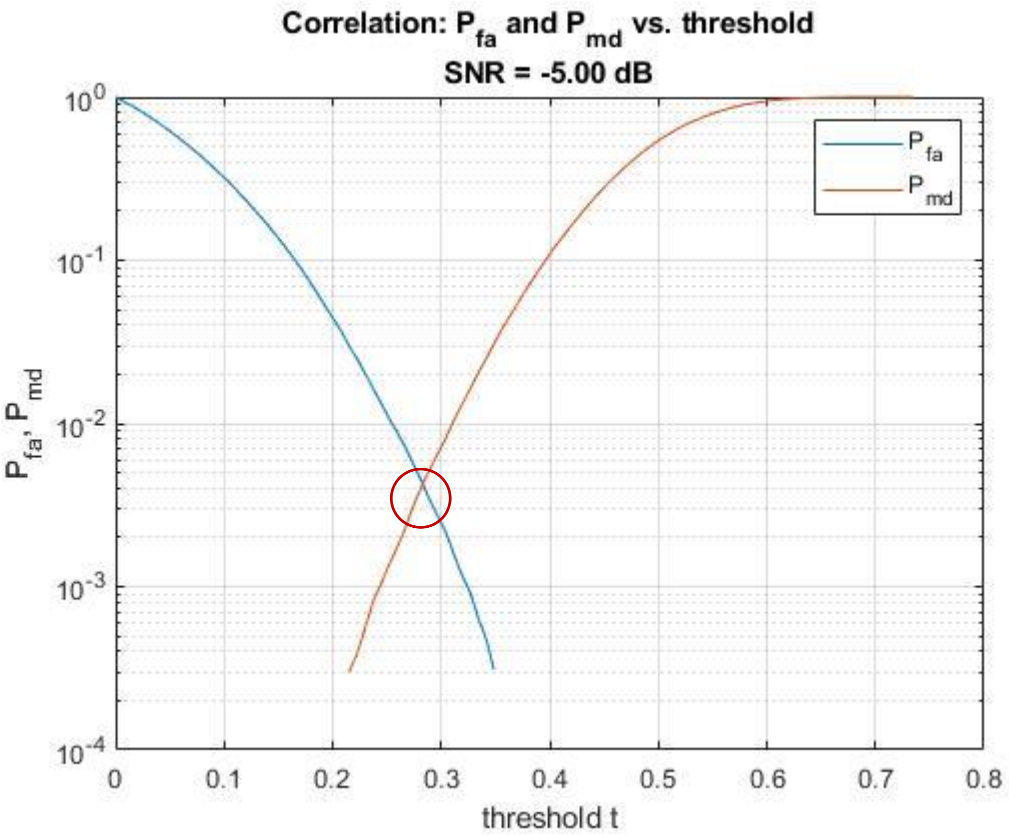

Figure 1.6.2: Plot of P_{fa} and P_{md} vs t curve with SNR = -5 dB

In this first example, a larger value of SNR than the previously studied one is chosen.
In figure 1.6.1 it can be seen that the ROC curve is almost a straight line, approaching $P_d = 1$ and $P_{fa} = 0$. As mentioned before, this behavior brings a satisfactory yield. Furthermore, in figure 1.6.2 we can observe that the probabilities P_{fa} and P_{md} are lower than in the previous example which once again is what we aim for.

ROC curve with different SNR values - example 2: smaller SNR

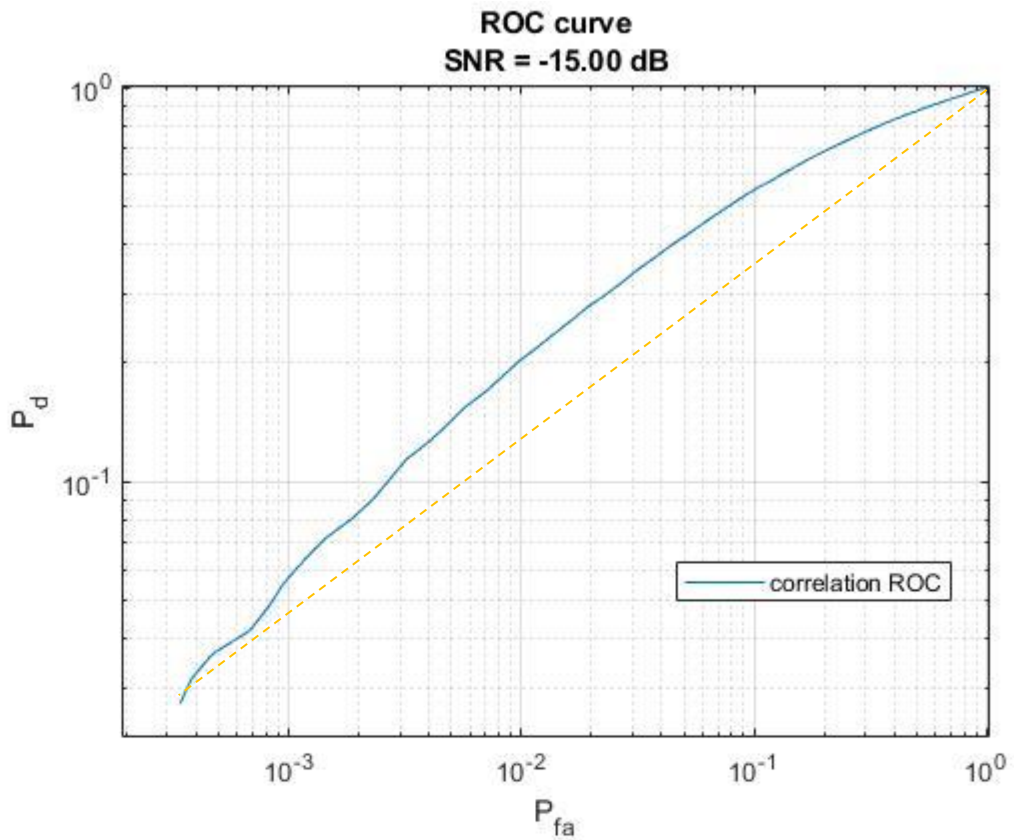

Figure 1.7.1: Plot of the ROC curve with SNR = -15 dB

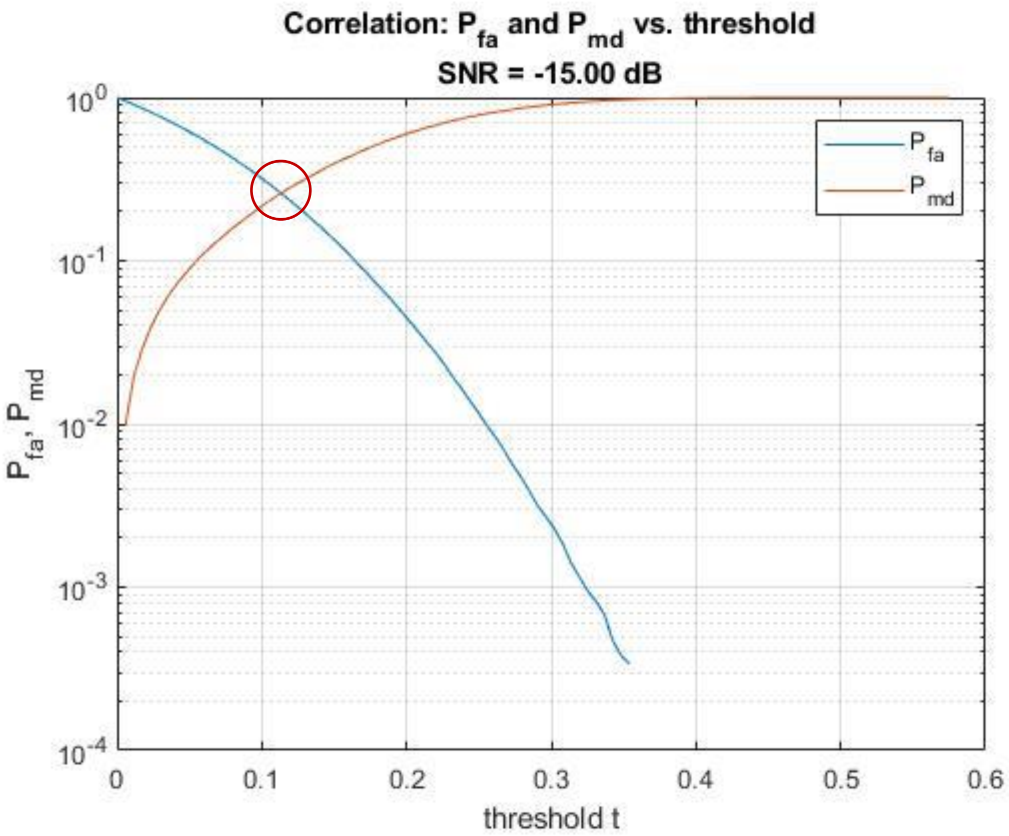

Figure 1.7.2: Plot of P_{fa} and P_{md} vs t curve with SNR = -15 dB

In the second example, a smaller value of SNR than the previously studied one is chosen.
In figure 1.7.1 the ROC curve is approaching a "diagonal. This behavior implies that the number of incorrect positive results approaches the number of correct positive results, which is not the preferred outcome.
In figure 1.7.2 it can be seen that the probabilities P_{fa} and P_{md} are quite high which once again is undesirable.