Report for Project 1

Jingxi Huang, Letao Feng

I. INTRODUCTION

In this project, we went through some fundamental concepts in elementary probability theory and signal processing. The problems are mainly about exploring some properties of the Gaussian distribution and parameter estimation, and analyzing system models with Gaussian noise. The solutions are carried out by MATLAB implications and mathematical derivations, which will be explained and discussed in this report.

II. PROBLEM FORMULATION AND SOLUTION

A. Task 1

First, to estimate the mean and the variance of the distribution from the sequences, we can use the estimators for the discrete data sequence:

$$\hat{m}_X = \frac{1}{N}\sum_{n=1}^{N} x(n) \tag{1}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N} (x(n) - \hat{m}_X)^2 \tag{2}$$

In MATLAB, we can use the functions mean() and var(x,1) to realize (1) and (2) respectively. After loading the data from Gaussian1D.mat, it's easy to calculate the mean and variance of x1(n), x2(n) and x3(n), which are 1.3829 and 5.6385, 0.6135 and 2.1711, 0.4408 and 1.9596. Then, we can use cdfplot() to plot the empirical distribution of each sequence, and compare the curves with the original Gaussian distribution, which is plotted by normcdf(). As shown in figure 1, it's clear to see that as the length of the sequence increases, the two curves become closer together and the approximation of the original distribution becomes more accurate.

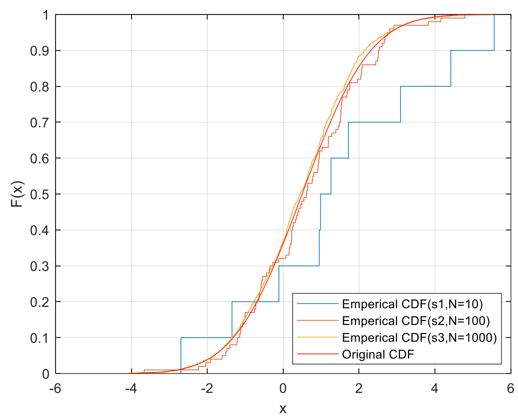


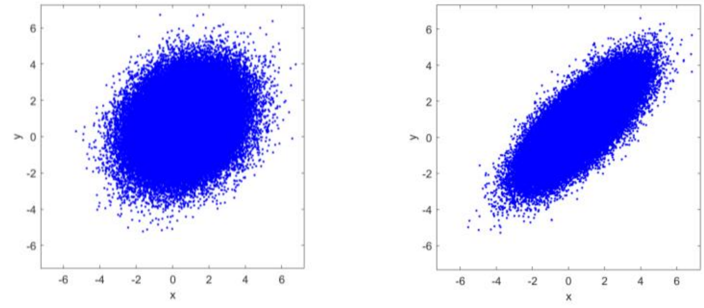
Figure 1: Empirical distribution

B. Task 2

The general expression for the joint Gaussian distribution is:

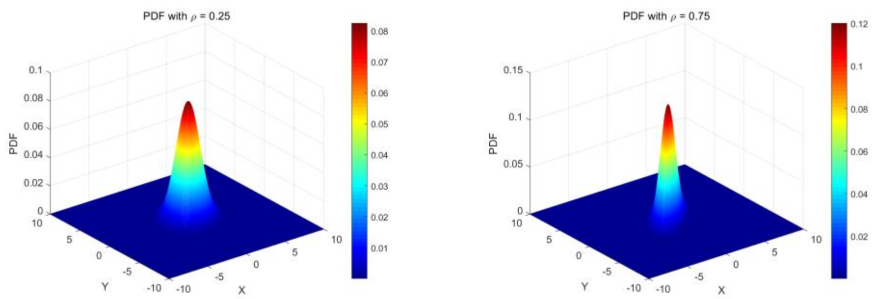
$$f_{XY}(x,y)=\left(2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}\right)^{-1}\exp\left[-\frac{1}{2\left(1-\rho^2\right)}\left(\frac{\left(x-m_X\right)^2}{\sigma_x^{\ 2}}-\frac{2\rho\left(x-m_X\right)\left(y-m_Y\right)}{\sigma_x\sigma_y}+\frac{\left(y-m_Y\right)^2}{\sigma_Y^{\ 2}}\right)\right] \tag{3}$$

To decide which value of ρ belongs to which sequence, we can first use the scatter plot of X and Y. From figure 2, it's easy to tell that the shape of $\{x_1(n), y_1(n)\}$ looks more like an ellipse than $\{x_2(n), y_2(n)\}$, which indicates a higher value of ρ.



(a) {x1(n), y1(n)}　　(b) {x2(n), y2(n)}

Figure 2: Scatter plot of X and Y.



(a) {x1(n), y1(n)}　　(b) {x2(n), y2(n)}

Figure 3: Empirical pdfs.

Then we can calculate the mean and standard deviation of X and Y, and calculate the pdf in (3) using the estimated parameters. At last, we can plot the pdf in 3D by using the function mesh(), as shown in figure 3.

If the pdf was characterized from −ρ, instead of ρ, it implies a negative linear correlation between the variables. Therefore the shape will be in the opposite direction along a different diagonal axis in xy-plane, forming an elliptical shape.

C. Task 3

From the given conditions, the expression of $f_Z(z)$ can be written as:

$$f_Z(z)=\frac{f_{XY}(x,y)}{f_Y(y)} \tag{4}$$

Since the expression of $f_{XY}(x,y)$ is given in the previous task, we only need to know the expression of $f_Y(y)$, that is:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{\left(y - m_y\right)^2}{2\sigma_y^{\ 2}} \right) \tag{5}$$

Then substitute (3), (5) and $\sigma x^2 = \sigma y^2 = \sigma^2$ in (4)， and the conditional pdf $f_Z(z)$ is obtained:

$$f_z(z) = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma} \exp\left(-\frac{\left[x - \rho y - \left(m_x - \rho m_y\right)\right]^2}{2\left(1-\rho^2\right)\sigma^2} \right) \tag{6}$$

Because X and Y are i.i.d Gaussian, we can derive that $m_{x+y} = m_x + m_y$, $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y$, $m_{x-y} = m_x - m_y$, $\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$. So the pdf of X+Y and X-Y can be derived from the expression of pdf in Gaussian distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left(x - m_x\right)^2}{2\sigma^2} \right] \tag{7}$$

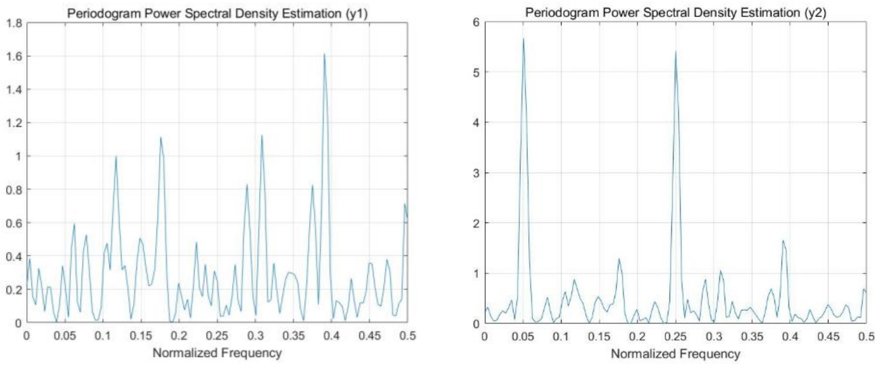
The results are as below:

$$f_{X+Y}(x+y) = \frac{1}{2\sqrt{\pi}\sqrt{2(1+\rho)}\sigma} \exp\left[-\frac{\left(x + y - m_x - m_y\right)^2}{4(1+\rho)\sigma^2} \right] \tag{8}$$

$$f_{X-Y}(x-y) = \frac{1}{2\sqrt{\pi}\sqrt{2(1-\rho)}\sigma} \exp\left[-\frac{\left(x - y - m_x + m_y\right)^2}{4(1-\rho)\sigma^2} \right] \tag{9}$$

D. Task 4

In Task 4, the two periodograms of the output sequences are shown below:



(a) (b)

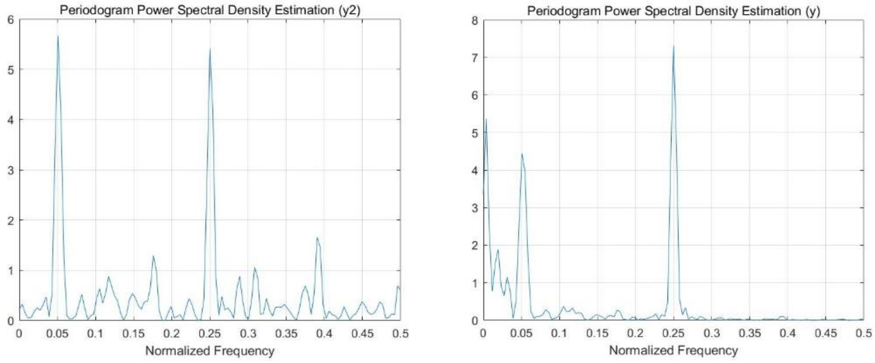
Figure 4: Periodogram Power Spectral Density(PSD) for y1 and y2.

In Figure 4(a), there are no special frequencies in it, that is to say, no obvious peak frequencies. This is the same as the properties of the white noise power spectral density. As a result, the y1 output sequence can be related to the case H0.

However, in Figure 4(b), there are two obvious peak frequencies. The first is in 0.05, and the second is in 0.25. These frequencies match the normalized frequencies $v_0 = 0.05$ and $v_1 = 0.25$ in the two sinusoidal signals. Therefore, the y2 output sequence can be related to case H1. According to these data, the sinusoidal frequencies are recovered with very high accuracy and almost completely.

E. Task 5

The periodogram is shown below:



(a) **(b)**

Figure 5: Periodogram Power Spectral Density(PSD) for y2 and y.

In Figure 5(b), there are also some obvious peak frequencies in it. However, compared to Figure 5(a), there is another peak frequency of 0.0039. This is not a normalized frequency in the sinusoidal signal, which means that in this situation, the accuracy of the recovered sinusoidal frequencies v_0 and v_1 is poor in the presence of coloured noise.

According to the definition of white noise, the noise correlation is equal to 0. This is not the case for coloured noise, which is correlated noise. Therefore, compared to Figure 5(a) and Figure 5(b), the conclusion is: the stronger the noise correlation, the worse the recovery effect.

F. Task 6

To solve $r_{x_1}(k)$, we set $k=0$ and $k>0$ respectively, and then we can get

$$r_{x_1}(0) = E[(\alpha x_1(n-1) + z(n))^2] = \alpha^2 r_{x_1}(0) + \sigma_z^2 \tag{10}$$

$$r_{x_1}(k) = E[(\alpha x_1(n-1) + z(n)) x_1(n-k)] = \alpha r_{x_1}(k-1) \tag{11}$$

As a result, in general we get

$$r_{x_1}(k) = \alpha^{|k|} \frac{\sigma_z^2}{1-\alpha^2} \tag{12}$$

Then, $R_{x_1}(v)$ can be obtained by doing a DTFT transform on $r_{x_1}(k)$.

$$R_{x_1}(v) = \frac{\sigma_z^2}{1+\alpha^2 - 2\alpha\cos(2\pi v)} = \frac{1}{1.0625 - 0.5\cos(2\pi v)} \tag{13}$$

For $R_{x2}(v)$, it can be calculated as $R_{x2}(v) = |H(v)|^2 \cdot R_{x1}(v)$. By taking the Fourier transform of $h(n)$, the final answer can be written as

$$R_{x_2}(v) = \frac{1}{1-2\beta\cos(2\pi v) + \beta^2} \cdot \frac{\sigma_z^2}{1+\alpha^2 - 2\alpha\cos(2\pi v)} = \frac{1}{(1.0625 - 0.5\cos(2\pi v))^2} \tag{14}$$

Two power spectra plots are shown below:

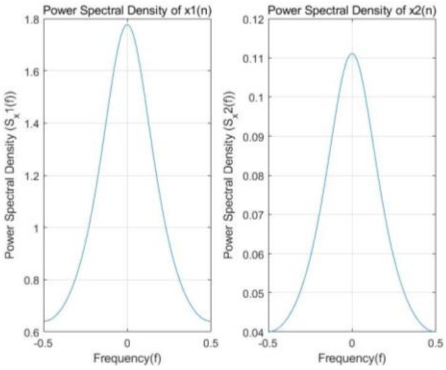


Figure 6: PSD for x_1(n) and x_2(n).

G. Task 7

Based on $R_{x2}(v)$, $r_{x2}(k)$ can be obtained by doing an inverse Fourier Transform.

$$
\begin{aligned}
r_{x_2}(k) &= \frac{1}{1+\alpha^2-2\alpha\cos(2\pi v)} * \frac{1}{1+\alpha^2-2\alpha\cos(2\pi v)} \\
&= 0.25^{|k|}\frac{1}{1-0.25^2} * 0.25^{|k|}\frac{1}{1-0.25^2} \\
&= 1.138\left[\left(\frac{1}{4}\right)^{|k|} * \left(\frac{1}{4}\right)^{|k|}\right] = 1.138\left[\sum_{m=-\infty}^{\infty}\left(\frac{1}{4}\right)^{|m|}\left(\frac{1}{4}\right)^{|k-m|}\right]
\end{aligned} \tag{15}
$$

In conclusion,

$$
r_{x_2}(k) = 1.138 \cdot \left(\frac{1}{4}\right)^{|k|}(|k|+1.133) \tag{16}
$$

Then we can draw the plot in MATLAB:

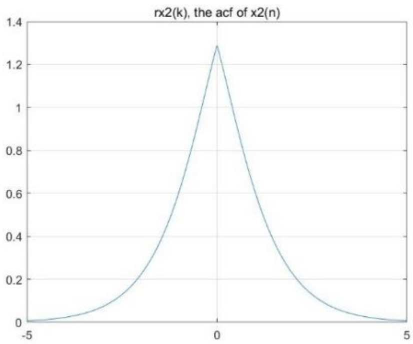


Figure 7: The acf for x_2(n).

III. CONCLUSIONS

This project starts with the study of one- and two-dimensional Gaussian distributions, followed by the study of system models with different kinds of Gaussian noise. Analyzing the power spectral density is crucial to study the results of the signal output. By solving different tasks, this paper provides insights into the properties associated with Gaussian noise and its behaviour in signal processing systems.