Report of Project 1, EQ1220

[redacted]

I. INTRODUCTION

In this project, the authors deploy the knowledge learnt from class and textbook [1] to solve practical problems in MATLAB. For the first three tasks, **Task 1-3**, we explore the properties of sequences that follow Gaussian distribution by plotting the corresponding empirical distribution and explore how the parameters will affect Gaussian distribution. In **Task 4** and **5**, we intend to figure out to what extent the different types of noises will interfere the transmitted sinusoidal signal from the periodograms, which are estimators of the power spectral density (PSD) of signals [2]. In **Task 6** and **7**, we focus on two important concepts in this course: auto-regressive (AR) processes and linear time-invariant (LTI) systems. We characterize an LTI system whose input is an autoregressive signal, including autocorrelation functions and PSD. The authors intend to consolidate the knowledge of probability theory and stochastic signal analysis, and improve problem-solving skills using programming (MATLAB), through conducting this project.

II. PROBLEM FORMULATION AND SOLUTION

Task 1 (The Gaussian Distribution)

There are three sequences $\{x_i(n)\}_{i=1}^{3}$ with different length N_i sampled from a Gaussian distribution according to $\mathcal{N}\sim(0.5,2)$. Firstly, we estimate the mean and variance of each sampled sequence based on the following equations respectively. The results are shown in Table 1.

$$\widehat{m_x} = \frac{1}{N}\sum_{n=0}^{N-1} x(n), \qquad \widehat{Var}[x] = \frac{1}{N-1}\sum_{n=0}^{N-1}(x(n)-\hat{m}_x)^2 .$$

After that the empirical distribution of each sequence is calculated by histogram function and illustrated in Fig 1. In this case, each bin will cover a range of 0.3 units along the x-axis and the histogram is normalized to represent probabilities. In other words, the heights of the histogram bars will represent the probability density of the data in each bin, and the area under the histogram will sum to 1. Additionally, the pdf of original gaussian distribution without sampling is presented in Fig 2.

	Mean	Variance
$x_1(n)$	1.38	6.27
$x_2(n)$	0.61	2.19
$x_3(n)$	0.44	1.96

Table. 1. The estimated mean and variance of $\{x_i(n)\}_{i=1}^{3}$

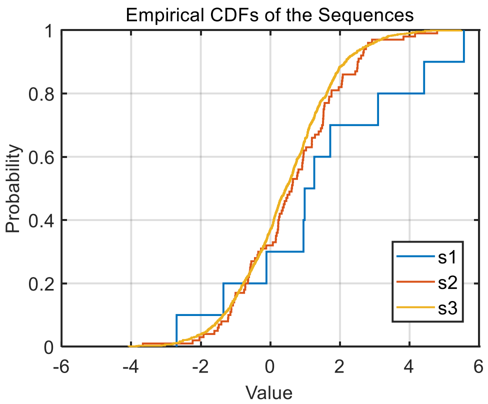


Fig. 1. The probability distribution of $\{x_i(n)\}_{i=1}^{3}$.

According to the results, larger sequence lengths N_i generally lead to more accurate and reliable estimations of the original distribution. As N_i increases, the estimates become more robust, with reduced sensitivity to sample fluctuations, resulting in more accurate representations of the original distribution.

Task 2 (The Gaussian Distribution)

The general expression of the joint Gaussian distribution is expressed in formula (3). In the formula σ_x, σ_y mean the standard deviation of X and Y respectively. m_x/σ_x and m_y/σ_x are the expectations / standard deviations of X and Y, while ρ is the correlation coefficient between X and Y. The 3D-plots of empirical pdfs are illustrated in Fig 2 with corresponding values of ρ. The correlation coefficient of s1 is larger than that of s2 since. Since the probability density function pdf of s2 approximating more closely to Y=X, the correlation coefficient of s2 is consequently higher. In general, the magnitude of ρ influences the degree of elongation or circularity, with larger magnitudes indicating stronger linear relationship and the sign of ρ determines the orientation of the elliptical shape of the joint PDF. Positive ρ leads to an elongation in one direction, negative ρ leads to an elongation in the opposite direction, and zero ρ results in a circular shape.

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} + \frac{\left(y-m_y\right)^2}{\sigma_y^2} \pm 2\rho\frac{(x-m_x)\left(y-m_y\right)}{\sigma_x\sigma_y}\right]\right). \quad (1)$$

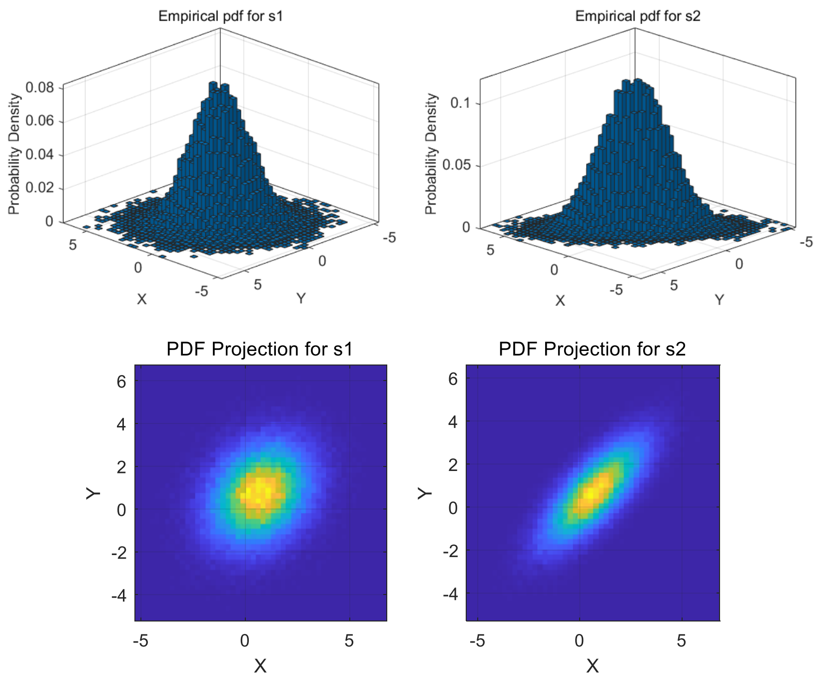


Fig. 2. The probability density function of sequence 1 and 2 with $\rho_1 = 0.25$, ρ_2 =0.75.

Task 3 (The Gaussian Distribution)

Knowing the means and variances of Gaussian distribution of X and Y, $f_Z(z), Z = X|Y = y$ can be derived from

$$f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} exp\left(-\frac{(x-m_x)^2 - 2\rho(x-m_x)(y-m_y) + \rho^2\left(y-m_y\right)^2}{2\sigma^2(1-\rho^2)}\right),$$

replacing x with z, we get

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} exp\left(-\frac{\left(z-\left(m_x+\rho\left(y-m_y\right)\right)\right)^2}{2\sigma^2(1-\rho^2)}\right). \quad (2)$$

If X and Y are Gaussian distributions with correlation coefficient ρ, $X+Y$ and $X-Y$ also follows normal distribution. Assume $U=X+Y$, it is easy to prove that $m_u = m_x + m_y$ and $\sigma_u^2 = Var(U) = Var(X+Y) = E\left[\left(X+Y-m_x-m_y\right)^2\right]$

$$= E\left[(X-m_x)^2 + \left(Y-m_y\right)^2 + 2(X-m_x)\left(Y-m_y\right)\right] = Var(X) + Var(Y) + 2Cov(X,Y)$$

$$= \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y$$

So, the pdf for $U = X + Y$ is

$$f_U(u) = \frac{1}{\sqrt{2\pi{\sigma_u}^2}} exp\left(-\frac{(u-m_u)^2}{2{\sigma_u}^2}\right). \tag{3}$$

Similarly, the pdf for $V = X - Y$ is

$$f_V(v) = \frac{1}{\sqrt{2\pi{\sigma_v}^2}} exp\left(-\frac{(v-m_v)^2}{2{\sigma_v}^2}\right), \tag{4}$$

where $m_v = m_x - m_y$ and $\sigma_u^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$.

Task 4 (System Models with Gaussian Noise)

In this task, we intend to explore the behavior of the estimated PSDs, the periodograms, of two types of signals after they are interfered by Gaussian noise. The periodogram of output $y_0(n)$ and $y_1(n)$ are

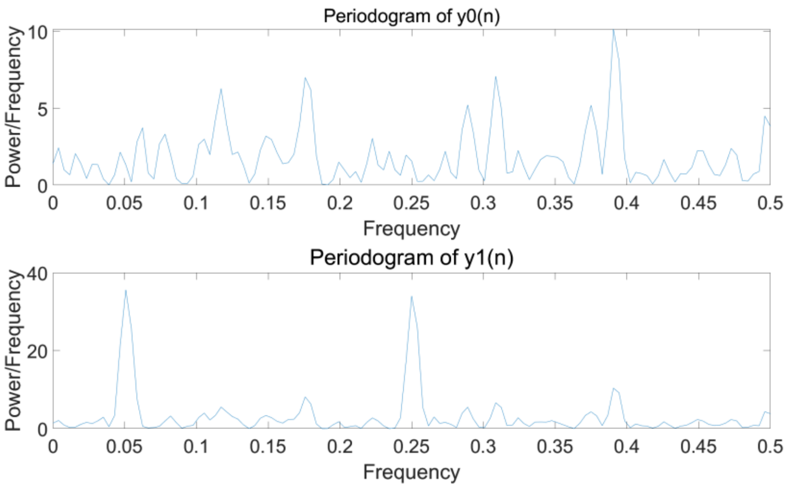
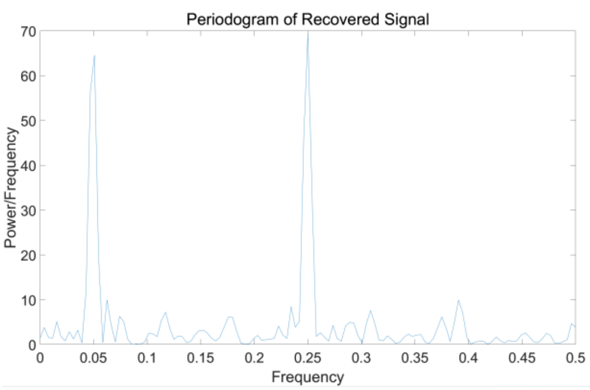


Fig. 3. Periodogram of sequence $y_0(n)$ (above) and $y_1(n)$ (below).



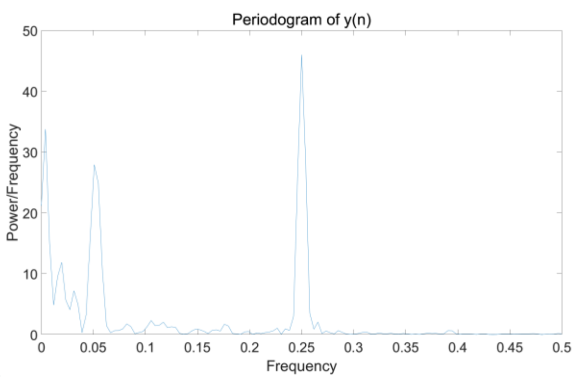


Fig. 4. Periodogram of Recovered Signal (left).

Fig. 5. Periodogram of Signal $y(n)$ with correlated noise samples (right).

presented in Fig 3. From the result, the figure above shows the periodogram in case H0 (a white noise sequence) and figure below present the periodogram in case H1 (a sinusoidal signal).

Task 5 (System Models with Gaussian Noise)

We would like to further explore the behavior of periodogram after a sinusoidal signal is corrupted by colored noise. Fig 4 presents the periodogram of recovered signal when subtracting white noise. The highest values are clearly observed at frequencies 0.05 and 0.25. The Fig 5 illustrates the periodogram of signal distorted by colored noise signal. White noise exhibits uniform power spectral density across all frequencies, thereby avoiding frequency-related biases during signal restoration. Conversely, colored noise introduces frequency-dependent characteristics that can complicate the restoration process.

Task 6 (System Models with Gaussian Noise)

In signal analysis, power spectral density is an essential tool which represent how the energy of a signal is distributed in the frequency domain [3]. Consider passing an AR(1) process $x_1(n) = \alpha x_1(n-1) + z(n)$ through an LTI system with impulse response $h_2(n) = \beta^n u(n)$, where $z(n)$ is white noise with variance $\sigma_z^2 = 1$. $\alpha = \beta = 0.25$ and $u(n)$ is the unit step function. The output of the system is $x_2(n) = \sum_{k=0}^{+\infty} h_2(k) x_1(n-k)$. To derive the power spectrum of $x_1(n)$, $R_{x_1}(v)$, first derive the autocorrelation function $r_{x_1}(n)$ of $x_1(n)$. Refer to page 65 of textbook [1], it is easy to get $r_{x_1}(0) = \frac{{\sigma_z}^2}{1-\alpha^2}$, and

$$r_{x_1}(k) = E\left[\left(\alpha x_1(n-1) + z(n)\right) x_1(n-k)\right] = \alpha r_{x_1}(k-1) = a^{|k|} \frac{{\sigma_z}^2}{1-\alpha^2}. \tag{5}$$

Then computing its power spectrum by DTFT, refer to page 108 of textbook [1]

$$R_{x_1}(v) = \mathcal{F}_d\left[r_{x_1}(k)\right] = \frac{{\sigma_z}^2}{1+\alpha^2 - 2\alpha\cos(2\pi v)} = \frac{16}{17 - 8\cos{(2\pi v)}}. \tag{6}$$

Knowing that $H(v) = \mathcal{F}_d[h(t)] = \frac{1}{1-\beta e^{-j2\pi v}}$, the power spectrum of $x_2(n)$, $R_{x_2}(v)$ is

$$R_{x_2}(v) = R_{x_1}(v)|H(v)|^2 = \frac{16}{17 - 8\cos{(2\pi v)}} \times \frac{16}{17 - 8\cos{(2\pi v)}} = \left(\frac{16}{17 - 8\cos(2\pi v)}\right)^2. \tag{7}$$

Fig 6 demonstrates the plots of $R_{x_1}(v)$ and $R_{x_2}(v)$.

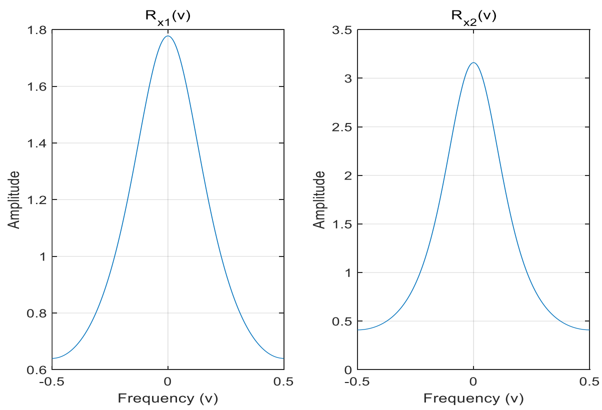


Fig. 5. Plots of Power Spectrum of $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$

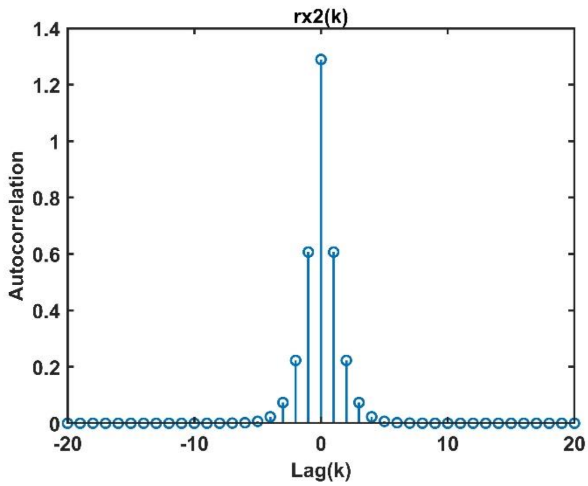


Fig. 6. The plot of acf of $\boldsymbol{x_2}$, $\boldsymbol{r_{x_2}(k)}$

Task 7 (System Models with Gaussian Noise)

In signal analysis, autocorrelation function quantifies the degree of similarity between the signal and a shifted version of itself over varying time lags [3]. We have known the expression for $r_{x_1}(k)$, the *acf* of $x_2(n)$, $r_{x_2}(k)$ can be calculated via

$$r_{x_2}(k) = \sum_{u=-\infty}^{+\infty} \sum_{v=-\infty}^{+\infty} h_2(u) h_2(v) r_{x_1}(k-u+v) = r_{x_1}(k) * h_2(k) * h_2(-k). \tag{8}$$

By formula (10), we can plot $r_{x_2}(k)$ using $r_{x_1}(k)$ and $h_2(k)$, see Fig 8. Since $R_{x_1}(v) = |H(v)|^2$, $R_{x_2}(k)$ can be calculated via $R_{x_2}(v) = R_{x_1}(v) R_{x_1}(v)$. Then $r_{x_2}(k)$ can be derived from:

$$r_{x_2}(k) = r_{x_1}(k) * r_{x_1}(k) = \left(\frac{16}{15}\right)^2 \sum_{n=-\infty}^{+\infty} (\frac{1}{4})^{|n|} \cdot \left(\frac{1}{4}\right)^{|k-n|}. \tag{9}$$

When $k \geq 0$: $r_{x_2}(k) = \left(\frac{16}{15}\right)^2 \sum_{n=0}^{k} \left(\frac{1}{4}\right)^n \cdot \left(\frac{1}{4}\right)^{k-n} + \sum_{n=k+1}^{+\infty} \left(\frac{1}{4}\right)^n \cdot \left(\frac{1}{4}\right)^{n-k} + \sum_{n=-\infty}^{-1} \left(\frac{1}{4}\right)^{-n} \cdot \left(\frac{1}{4}\right)^{k-n}$

$$= \left(\frac{16}{15}\right)^2 \left[(n+1) \cdot \left(\frac{1}{4}\right)^k + \frac{16}{15} \cdot \left(\frac{1}{4}\right)^{k+2} + \frac{1}{15} \cdot \left(\frac{1}{4}\right)^k\right] = \frac{256}{225} \times (k + \frac{17}{15}) \times (\frac{1}{4})^k. \tag{10}$$

When $k < 0$: $r_{x_2}(k) = \frac{256}{225} \times (-k + \frac{17}{15}) \times (\frac{1}{4})^{-k}$, the calculation is similar to above.

Therefore:

$$r_{x_2}(k) = \frac{256}{225} \times (|k| + \frac{17}{15}) \times (\frac{1}{4})^{|k|}. \tag{11}$$

III. CONCLUSIONS

In conclusion, we have successfully delved into various aspects of probability theory and stochastic signal analysis through a series of tasks. The initial tasks, Task 1-3, allowed us to delve into the characteristics of sequences following Gaussian distribution N, where we visualized empirical distributions and analyzed how different length of sequence influenced the Gaussian distribution. Moving on to Task 4 and 5, we investigated the impact of various types of noise on transmitted sinusoidal signals using periodograms as estimators of the power spectral density (PSD). In Task 6 and 7, we honed our understanding of auto-regressive (AR) processes and linear time-invariant (LTI) systems. We not only derived key parameters but also created visual representations of LTI systems with autoregressive inputs, including autocorrelation functions and PSD.

IV. REFERENCES

[1] P. Handel, R. Ottoson, H. Hjalmarsson, *Signal Theory*, ch.1–8. KTH, 2012.

[2] “Periodogram - MATLAB & Simulink - MathWorks,” MathWorks.

[3] X. Li, *Random Signal Analysis*, 5th ed. Beijing: Publishing House of Electronics Industry, 2018.

[4] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, 4th ed. Boston: Addison-Wesley, 2012.