Signal Processing and Statistics
Final Exam
2021-2022
Time limit: 90 Minutes

This exam contains 3 pages (including this cover page) and 4 exercises.

---

1. **Fourier series**

   Let us consider a real-valued function $x(t)$ such that $x(t+2\pi) = x(t)$. Its Fourier series representation is given by

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t)\right).$$

   The coefficients a_0, a_n and b_n are given by

$$\begin{aligned} a_0 &= \frac{1}{\pi} \int_{-\pi}^{\pi} x(t)\ \mathrm{d}t, \\ a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} x(t) \cos(2\pi n f_0 t)\ \mathrm{d}t, \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} x(t) \sin(2\pi n f_0 t)\ \mathrm{d}t, \end{aligned}$$

   with $f_0 = {}^1\!/_T$ its fundamental frequency.

   (a) (2 points) Prove that the basis of real harmonic oscillations

$$\sin(2\pi n f_0 t), \quad \cos(2\pi n f_0 t), \quad n = 1, 2, \cdots$$

   forms an orthogonal basis, i.e. their inner product is equal to 0 if $m \neq n$ and non-zero otherwise.

   (b) (1 point) Using the results of the previous question, find formulas for the amplitudes c_n and phases θ_n in the expansion of the periodic signal $x(t)$ in terms of only cosines, i.e.

$$x(t) = \sum_{n=0}^{\infty} c_m \cos(2\pi n f_0 t + \theta_n).$$

   (c) (2 points) Using these results, show that

$$\frac{a_0^2}{2} + \sum_{n=1}^{\infty} \left(a_n^2 + b_n^2\right) = \frac{1}{\pi} \int_{-\pi}^{\pi} x^2(t)\ \mathrm{d}t.$$

   This result is known as *Parseval's theorem*.

2. **Stochastic systems**

   Consider the following continuous-time linear system

   $$\dot{x} = -\alpha x + w$$

   with $x \in \mathbb{R}$ the state variable, $w \in \mathbb{R}$ the noise, $\alpha > 0$ and $\sigma > 0$.

   (a) (1 point) Let us assume the autocorrelation function of the noise w is given by

   $$R_{ww}(\tau) = \sigma^2 \delta(\tau).$$

   What does this tell you about the properties of the noise process w ?

   (b) (2 points) The impulse response of $\dot{x} = \alpha x$ is given by

   $$h(\tau) = \begin{cases} e^{-\alpha\tau} & \text{if } \tau \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

   The cross-correlation function between $w(t)$ and $x(t)$ is given by

   $$\begin{aligned} R_{wx}(\tau) &= h(\tau) * R_{ww}(\tau) \\ &= \int_{-\infty}^{\infty} h(t) R_{ww}(\tau - t)\ \mathrm{d}t. \end{aligned}$$

   Give the analytical expression of $R_{wx}(\tau)$.

   (c) (2 points) The auto-correlation of $x(t)$ is given by

   $$R_{xx}(\tau) = h(-\tau) * h(\tau) * R_{ww}(\tau),$$

   Given the analytical expression of $R_{xx}(\tau)$. What is the typical time-scale over which the signal $x(t)$ is correlated ?

3. **Bivariate statistics**

   The value of 2 statistical variables X and Y is given in table 1 for 5 people.

	X	Y
Inividual 1	3	12
Inividual 2	4	14
Inividual 3	2	8
Inividual 4	5	19
Inividual 5	3	11

Table 1: Statistical variables X and Y evaluated for 5 people.

   (a) (1 point) Compute the marginal arithmetic means $\overline{X}$ and $\overline{Y}$ for each variable.

   (b) (1 point) Compute the marginal standard deviations $\sigma(X)$ and $\sigma(Y)$ for each variable.

(c) (1 point) Compute the covariance $\text{cov}(X, Y)$ between X and Y .

(d) (1 point) Suppose that a linear correlation holds between X and Y. Compute the equation of the regression line, X being the explanatory variable.

(e) (1 point) Compute the correlation coefficient between X and Y. What do you conclude?

4. **Principal component analysis**

Consider the table 2, in which the size and weight of 6 people is given. In this exercise, we apply the Principal Component Analysis (PCA) on this set of data.

	size [cm]	weight [kg]
Inividual 1	51	162
Inividual 2	64	165
Inividual 3	60	150
Inividual 4	90	190
Inividual 5	95	180
Inividual 6	85	185

Table 2: Size and weight of 6 people.

We note n the number of individuals and p the number of variables. We note $X \in M_{n,p}(\mathbb{R})$, the data matrix gathering the data of table 2 (each line of X corresponds to a specific individual, the first column of X corresponds to the variable "size" and the second one corresponds to the variable "weight").

(a) (1 point) Compute the line vectors $\overline{X}^T$, $\text{Var}(X)^T$ and $\sigma(X)^T$ gathering respectively the marginal arithmetic means, variances and standard deviations of each variable.

(b) (1 point) We note m the number of principal components that can be computed. What is the value of m? How many principal components must be considered to describe 100% of the variability of the data.

(c) (1 point) Compute the centered data matrix X_c and the covariance matrix of the centered data $\text{Cov}(X_c)$.

(d) (1 point) We note q_i $(1 \leq i \leq m)$ the m principal components. Compute the m principal components. For each value of i, compute the part (in %) of the variability of the data explained by the first i principal components.

(e) (1 point) Express the centered data matrix Y of the data coordinates in the principal components coordinate system.