Statistical Signal Processing and Applications: exercise book

Andrea Ridolfi
Luciano Sbaiz
Martin Vetterli
EPFL - I&C - LCAV

Chapter 1

Review Material

1.1 Linear Algebra

Exercise 1. Matrix properties

(a) Consider the following matrix:

$$\mathbf{X} = \begin{bmatrix} \frac{5}{4} & \frac{-\sqrt{3}}{4} \\ \frac{-\sqrt{3}}{4} & \frac{7}{4} \end{bmatrix}$$

What are the eigenvalues of $\mathbf{X}$? Is it possible to diagonalize the matrix? Can $\mathbf{X}$ be the correlation of a random vector of length 2?

(b) What are the eigenvectors of the matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \\ -\sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix}?$$

Suppose a real symmetric matrix $\mathbf{B}$ has eigenvalues equal to 5 and 3. Given the matrix $\mathbf{A}$ above, what is the determinant of $\mathbf{AB}$?

(c) Consider the following matrix:

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 & 2 \\ 2 & 1 & 2 & 3 \\ 3 & 2 & 1 & 2 \\ 2 & 3 & 2 & 1 \end{bmatrix}$$

What type of matrix is this? How would you diagonalize it without computing the eigenvalues and eigenvectors?

Solution 1. Matrix properties

(a) The eigenvalues are obtained by solving the equation $\det(\mathbf{X} - \lambda\mathbf{I}) = 0$, which is true for $\lambda = 1$ and $\lambda = 2$.

If all eigenvalues are different, which is the case, then $\mathbf{X}$ can be diagonalized (although the opposite is not necessarily true!) into a matrix $\mathbf{D}$ such that $\mathbf{D} = \mathbf{V}^{-1}\mathbf{X}\mathbf{V}$, where $\mathbf{V}$ is a matrix with the eigenvectors as columns.

We may also notice that $\mathbf{X} = \mathbf{X}^T$, which is a sufficient condition for $\mathbf{X}$ to be diagonalizable.

If all the eigenvalues of $\mathbf{X}$ are non-negative, which is the case, in addition to being a symmetric matrix, then $\mathbf{X}$ can represent a correlation matrix of a random vector of length 2. A matrix of this type is called positive semidefinite.

(b) If $\mathbf{V}$ is a matrix with the eigenvectors of $\mathbf{A}$ as columns, and $\mathbf{D}$ is a diagonal matrix with the eigenvalues of $\mathbf{A}$ in the diagonal, then $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$. We then notice that

$$\mathbf{D} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \mathbf{V} = \begin{bmatrix} \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \\ -\sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix}$$

and thus $\mathbf{v}_0 = \begin{bmatrix} \cos\frac{\pi}{4} \\ -\sin\frac{\pi}{4} \end{bmatrix}$ and $\mathbf{v}_1 = \begin{bmatrix} \sin\frac{\pi}{4} \\ \cos\frac{\pi}{4} \end{bmatrix}$ are the eigenvectors of $\mathbf{A}$.

Real symmetric matrices are diagonalizable by orthogonal matrices, and thus $\mathbf{B} = \mathbf{T}\mathbf{G}\mathbf{T}^{-1}$ where $\mathbf{T}$ is orthogonal and $\mathbf{G}$ is a diagonal matrix with the eigenvalues 5 and 3. In the case of $\mathbf{A}$, the matrix $\mathbf{V}$ is also orthogonal. Thus, applying the determinant property $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$, we get

$$\begin{aligned}
\det(\mathbf{AB}) &= \det(\mathbf{A})\det(\mathbf{B}) \\
&= \det(\mathbf{V}^T)\det(\mathbf{D})\det(\mathbf{V})\det(\mathbf{T}^T)\det(\mathbf{G})\det(\mathbf{T}) \\
&= \det(\mathbf{V}^T\mathbf{V})\det(\mathbf{D})\det(\mathbf{T}^T\mathbf{T})\det(\mathbf{G}) \\
&= \det(\mathbf{I})\det(\mathbf{D})\det(\mathbf{I})\det(\mathbf{G}) \\
&= \det(\mathbf{D})\det(\mathbf{G}) \\
&= (2 \cdot 1) \cdot (5 \cdot 3) = 30
\end{aligned}$$

(c) Since $\mathbf{X}$ is a real symmetric matrix, it is diagonalized by an orthogonal matrix. However, the matrix is also circulant, and thus $\mathbf{X}$ is diagonalized by the DFT matrix (in this case, of size 4).

1.2 Hilbert Spaces

Exercise 2. Parseval's equality.
Given a finite dimensional space $W = \mathbb{R}^N$ and an orthonormal basis $\{\boldsymbol{v}^{(i)}\}$, $i = 0, \ldots, N-1$, verify that for any $\boldsymbol{x} \in W$

$$\|\boldsymbol{x}\|_2^2 = \sum_{i=0}^{N-1} \left| < \boldsymbol{x}, \boldsymbol{v}^{(i)} > \right|^2 .$$

Solution 2. Parseval's inequality
Let $\boldsymbol{x} \in W$, then $\boldsymbol{x} = \sum_{i=0}^{N-1} \alpha_i \boldsymbol{v}^{(i)}$, $\alpha_i =< \boldsymbol{x}, \boldsymbol{v}^{(i)} >$. Therefore

$$\|\boldsymbol{x}\|_2^2 =< \boldsymbol{x}, \boldsymbol{x} >=< \boldsymbol{x}, \textstyle\sum_{i=0}^{N-1} \alpha_i \boldsymbol{v}^{(i)} >= \sum_{i=0}^{N-1} \alpha_i^\star < \boldsymbol{x}, \boldsymbol{v}^{(i)} >,$$

Now, changing $\alpha_i =< \boldsymbol{x}, \boldsymbol{v}^{(i)} >$, we have

$$\|\boldsymbol{x}\|_2^2 = \textstyle\sum_{i=0}^{N-1} < \boldsymbol{x}, \boldsymbol{v}^{(i)} >^\star < \boldsymbol{x}, \boldsymbol{v}^{(i)} >= \sum_{i=0}^{N-1} \left| < \boldsymbol{x}, \boldsymbol{v}^{(i)} > \right|^2 .$$

Exercise 3. OPTIMAL APPROXIMATION.
Consider a Hilbert space H and a subspace W spanned by an orthonormal basis $\{\boldsymbol{v}^{(i)}\}$, $i \in I$ (one can assume a finite dimensional case for simplicity). Prove that the approximation $\widehat{\boldsymbol{x}} \in W$ of $\boldsymbol{x} \in H$ given by:

$$\widehat{\boldsymbol{x}} = \sum_{i \in I} < \boldsymbol{x}, \boldsymbol{v}^{(i)} > \boldsymbol{v}^{(i)}$$

satisfies:

(a) $(\boldsymbol{x} - \widehat{\boldsymbol{x}}) \perp \widehat{\boldsymbol{x}}$;

(b) $\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2$ is minimum among all linear approximations in W.

Solution 3. OPTIMAL APPROXIMATION
(a) From the property of the orthogonality we know that

$$(\boldsymbol{x} - \hat{\boldsymbol{x}}) \perp \hat{\boldsymbol{x}} \quad \Leftrightarrow \quad < \boldsymbol{x} - \hat{\boldsymbol{x}}, \hat{\boldsymbol{x}} >= 0,$$

Proof by writing out:

$$\begin{aligned}
< \boldsymbol{x} - \hat{\boldsymbol{x}}, \hat{\boldsymbol{x}} > &= < \boldsymbol{x}, \hat{\boldsymbol{x}} > - < \hat{\boldsymbol{x}}, \hat{\boldsymbol{x}} > \\
&= < \boldsymbol{x}, \sum_{i \in I} < \boldsymbol{x}, \boldsymbol{v}^{(i)} > \boldsymbol{v}^{(i)} > - \|\hat{\boldsymbol{x}}\|^2 \\
&= \sum_{i \in I} < \boldsymbol{x}, \boldsymbol{v}^{(i)} >^{\star} < \boldsymbol{x}, \boldsymbol{v}^{(i)} > - \|\hat{\boldsymbol{x}}\|^2 \\
&= \sum_{i \in I} \left| < \boldsymbol{x}, \boldsymbol{v}^{(i)} > \right|^2 - \|\hat{\boldsymbol{x}}\|^2 \\
&= \|\hat{\boldsymbol{x}}\|^2 - \|\hat{\boldsymbol{x}}\|^2 = 0
\end{aligned}$$

where the last equation holds from the Parseval's equality. Note that $\sum_{i \in I} \left| < \boldsymbol{x}, \boldsymbol{v}^{(i)} > \right|^2$ equals $\|\hat{\boldsymbol{x}}\|^2$ and not $\|\boldsymbol{x}\|^2$, because it represents the projection of $\boldsymbol{x}$ onto the orthonormal basis $\{\boldsymbol{v}^{(i)}\}$, which has less elements then those that would be required to fully represent $\boldsymbol{x}$.

(b) $\hat{\boldsymbol{x}}$ is a linear combination of the basis approximation $\hat{\boldsymbol{x}} = \sum_{n \in I} \beta_n \boldsymbol{v}^{(n)}$ for some β_n. For simplicity, we can assume that $\beta_n \in \mathbb{R}$. The case where $\beta_n \in \mathbb{C}$ can be analyzed in a similar way considering the real and the imaginary part of β_n.
We want to find the coefficients β_n for which we have the minimum of $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2$. Therefore,

$$\frac{\partial}{\partial \beta_i} \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2 = 0$$

Thus,

$$
\begin{aligned}
\frac{\partial}{\partial \beta_i} \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2 &= \frac{\partial}{\partial \beta_i} < \boldsymbol{x} - \hat{\boldsymbol{x}}, \boldsymbol{x} - \hat{\boldsymbol{x}} > \\
&= \frac{\partial}{\partial \beta_i} \left(< \boldsymbol{x}, \boldsymbol{x} > - < \boldsymbol{x}, \hat{\boldsymbol{x}} > - < \hat{\boldsymbol{x}}, \boldsymbol{x} > + < \hat{\boldsymbol{x}}, \hat{\boldsymbol{x}} >\right) \\
&= \frac{\partial}{\partial \beta_i} \Bigg(< \boldsymbol{x}, \boldsymbol{x} > - \sum_{n \in I} \beta_n < \boldsymbol{x}, \boldsymbol{v}^{(n)} > - \sum_{m \in I} \beta_m < \boldsymbol{v}^{(m)}, \boldsymbol{x} > \\
&\quad + \sum_{n \in I} \sum_{m \in I} \beta_n \beta_m < \boldsymbol{v}^{(n)}, \boldsymbol{v}^{(m)} > \Bigg) \\
&= \frac{\partial}{\partial \beta_i} \left(< \boldsymbol{x}, \boldsymbol{x} > - \sum_{n \in I} \beta_n < \boldsymbol{x}, \boldsymbol{v}^{(n)} > - \sum_{m \in I} \beta_m < \boldsymbol{v}^{(m)}, \boldsymbol{x} > + \sum_{n \in I} \beta_n^2 \right) \\
&= 0
\end{aligned}
$$

From which,

$$
- < \boldsymbol{x}, \boldsymbol{v}^{(i)} > - < \boldsymbol{v}^{(i)}, \boldsymbol{x} > + 2\beta_i = 0 \quad \Rightarrow \quad \beta_i = \mathrm{Re}\left\{ < \boldsymbol{x}, \boldsymbol{v}^{(i)} > \right\}
$$

as claimed.

Exercise 4. Hilbert Spaces in Probability.
Consider the random variables X_0, X_1, X_2 defined on the same probability space. Suppose that the mean of each variable is 0 and the joint correlation matrix is

$$\mathbf{R_X} = \mathbb{E}[[X_0 X_1 X_2]^T [X_0 X_1 X_2]] = \begin{bmatrix} 8 & 4 & 1 \\ 4 & 8 & 4 \\ 1 & 4 & 8 \end{bmatrix}.$$

Let us define Hilbert space H as the space generated by all the linear combinations of the variables X_0, X_1, and X_2, i.e.

$$H = \{a_0 X_0 + a_1 X_1 + a_2 X_2, a_0, a_1, a_2 \in \mathbb{R}\}.$$

(a) Determine an orthogonal basis, $\{Y_0, Y_1\}$ for the subspace W generated by X_0 and X_1.

(b) Find the best approximation of the variable X_2 in the subspace W, i.e. the random variable Y that minimizes $\mathbb{E}[|Y - X_2|^2]$, with $Y \in W$. (Hint: apply the projection theorem.)

Solution 4. Hilbert Spaces in Probability.
This exercise may seem strange at a first glance, but it is actually a standard exercise of linear algebra. One should just replace the scalar product used in $\mathbb{R}^N$ with

$$\langle X, Y \rangle = \mathbb{E}[XY],$$

where X and Y are random variables defined on the same probability space. One could easily verify that this product is actually a valid scalar product. The scalar product always induces a norm, defined by

$$\|X\| = \sqrt{\langle X, X \rangle}.$$

With these definitions, the space H is actually a Hilbert space. (One could verify that all the properties valid for vector spaces hold for the set H and also that H is complete.)
The space H is generated by the random variables X_0, X_1, X_2 which represent a basis of the space. They are the vectors of the space and one can apply the usual vector operations on them.

(a) The subspace W is the subspace of H generated by the vectors (i.e. the random variables) X_0 and X_1. To determine an orthogonal basis, one can apply the Gram-Schmidt procedure:

$$\begin{aligned} Y_0 &= \tfrac{X_0}{\|X_0\|} \\ Y_1 &= \tfrac{X_1 - \langle X_1, Y_0 \rangle Y_0}{\|X_1 - \langle X_1, Y_0 \rangle Y_0\|}, \end{aligned}$$

and replace the scalar product and the norm with the definitions that we presented earlier. We obtain,

$$\begin{aligned} Y_0 &= \tfrac{X_0}{2\sqrt{2}} \\ Y_1 &= \tfrac{X_1 - \langle X_1, \frac{X_0}{2\sqrt{2}} \rangle \frac{X_0}{2\sqrt{2}}}{\|(X_1 - \langle X_1, \frac{X_0}{2\sqrt{2}} \rangle \frac{X_0}{2\sqrt{2}}\|} \\ &= -\tfrac{1}{2\sqrt{6}} X_0 + \tfrac{1}{\sqrt{6}} X_1. \end{aligned}$$

(b) To determine the best approximation of X_2 in W, say $\hat{X}_2$, we write it as a linear combination of X_0 and X_1 (or equivalently of Y_0 and Y_1),

$$\hat{X}_2 = b_0 X_0 + b_1 X_1.$$

The error of the approximation is given by

$$E = X_2 - \hat{X}_2$$

To apply the projection theorem we impose that the approximation error is orthogonal to W. This correspond to the two equations:

$$\begin{aligned} \langle E, X_0 \rangle &= 0 \\ \langle E, X_1 \rangle &= 0, \end{aligned}$$

which gives the linear system:

$$\begin{cases} \langle X_0, X_0 \rangle b_0 + \langle X_1, X_0 \rangle b_1 &= \langle X_2, X_0 \rangle \\ \langle X_0, X_1 \rangle b_0 + \langle X_1, X_1 \rangle b_1 &= \langle X_2, X_1 \rangle. \end{cases}$$

The solution of the system is

$$\begin{aligned} b_0 &= -\tfrac{1}{6} \\ b_1 &= \tfrac{7}{12}; \end{aligned}$$

therefore, $\hat{X}_2 = -X_0/6 + 7X_1/12$.

Exercise 5. FOURIER BASIS.
Consider the *Fourier basis* $\{\boldsymbol{w}^{(k)}\}_{k=0,\dots,N-1}$, defined as:

$$\boldsymbol{w}_n^{(k)} = e^{-j\frac{2\pi}{N}nk}.$$

(a) Prove that it is an *orthogonal basis* in $\mathbb{C}^N$. The inner product is define as in $\boldsymbol{l_2}$ space.

(b) Normalize the vectors in order to get an *orthonormal basis.*

(c) Propose the best least square approximation $\hat{\boldsymbol{y}} \in \mathbb{C}^N$ of a general vector $\boldsymbol{y} \in \mathbb{C}^{N+1}$.

Solution 5. FOURIER BASIS
(a) Fourier basis is a sequence of N-dimensional vectors

$$\left\{\boldsymbol{w}^{(k)} = \left(w_0^{(k)}, w_1^{(k)}, \dots, w_{N-1}^{(k)}\right)\right\}_{k=0,\dots,N-1}.$$

Recall that the set of N non-zero orthogonal vectors in an N-dimensional subspace is a basis for the subspace. Therefore, we need to prove the orthogonality across the vectors $\left\{\boldsymbol{w}^{(k)}\right\}_{k=0,\dots,N-1}$. Let us compute the inner product, that is:

$$\begin{aligned} < \boldsymbol{w}^{(k)}, \boldsymbol{w}^{(h)} > &= \sum_{n=0}^{N-1} w_n^{(k)} w_n^{(h)\star} = \sum_{n=0}^{N-1} e^{-j\frac{2\pi}{N}nk} e^{j\frac{2\pi}{N}nh} \\ &= \sum_{n=0}^{N-1} e^{-j\frac{2\pi}{N}n(k-h)} = \begin{cases} N & if\, k=h \\ 0 & otherwise \end{cases} \end{aligned}$$

Since the inner product of the vectors is equal to 0 for $k \neq h$, we conclude that they are orthogonal. However, they do not have a unit norm and therefore are not the orthonormal vectors.

(b) In order to obtain the *orthonormal basis* we normalize the vectors with the factor $1/\sqrt{N}$, having:

$$\begin{aligned} < \boldsymbol{w}_{norm}^{(k)}, \boldsymbol{w}_{norm}^{(h)} > &= \sum_{n=0}^{N-1} \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}nk} \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}nh} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} e^{-j\frac{2\pi}{N}n(k-h)} = \begin{cases} 1 & if\, k=h \\ 0 & otherwise \end{cases} \end{aligned}$$

(c) In order to use the projection theorem we need to define a new space S that is a subspace of $\mathbb{C}^{N+1}$. A natural extension from $\mathbb{C}^N$ to $\mathbb{C}^{N+1}$ is to define $S = \left\{(c^T, 0)^T,\, c \in \mathbb{C}^N\right\}$. In that case, the orthonormal basis for the space S is $\boldsymbol{w}_s^{(k)} = \left(\boldsymbol{w}_{norm}^{(k)T}, 0\right)^T$. Now, the best linear approximation of $\boldsymbol{y} \in \mathbb{C}^{N+1}$ on the subspace S, which minimizes the norm $\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|$, is obtained by projecting $\boldsymbol{y}$ onto an orthogonal basis $\boldsymbol{w}_s^{(k)}$,

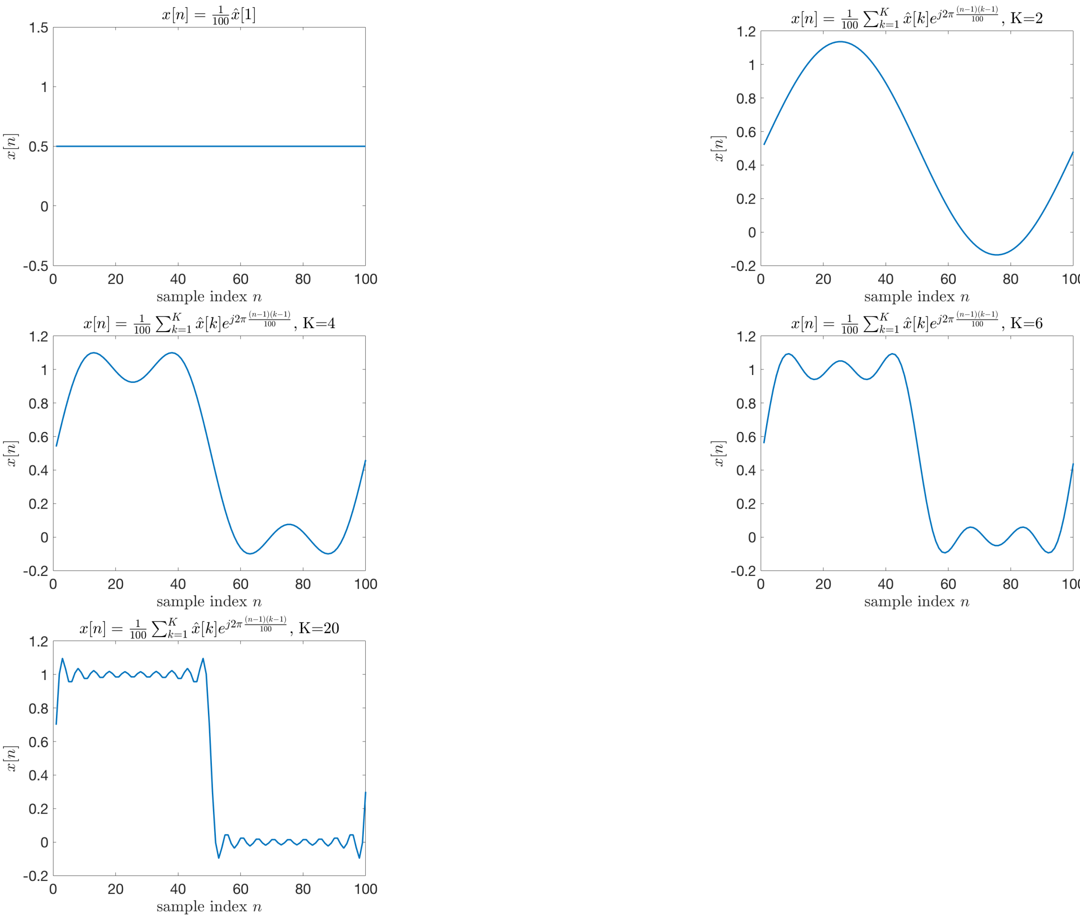
$$\hat{\boldsymbol{y}} = \sum_{k=0}^{N-1} < \boldsymbol{y}, \boldsymbol{w}_s^{(k)} > \boldsymbol{w}_s^{(k)}$$

Exercise 6. Coding: Fourier series series approximation
Using Matlab or Python (`numpy.fft`), plot a step function together with it approximation using k first elements of it's Fourier series. For example, you can take a length 100 sequence, where first 100 elements are equal to 1 and the rest is 0. Generate plots for different k. What do you observe?

Solution 6.
If you take signal of the length 100, then it can be decomposed as follows:

$$x[n] = \frac{1}{100}\sum_{k=1}^{100} \hat{x}[k] e^{j2\pi\frac{(n-1)(k-1)}{100}}\ ,\ n = 1\ldots 100$$

By truncating this series, we can approximate this series with different number of of sine waves, in the examples below $1, 2, 6$ and 20:



You can observe that even though the approximations gets closer and closer to the step function in general, there are regions where the difference stays large: the sine waves are not well suited for approximating discontinuities.

Exercise 7. CIRCULANT MATRICES.
Prove that the Fourier basis vectors $(1, e^{-j\frac{2\pi}{N}k}, \ldots, e^{-j\frac{2\pi}{N}(N-1)k})^T$ are eigenvectors of an $N \times N$ circulant matrix.

Solution 7. CIRCULANT MATRICES
Let's write the $N \times N$ circulant matrix as

$$\mathbf{C} = \begin{bmatrix} c_0 & c_1 & \cdots & c_{N-1} \\ c_{N-1} & c_0 & \cdots & c_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & \cdots & c_0 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0^T \\ \mathbf{c}_1^T \\ \cdots \\ \mathbf{c}_{N-1}^T \end{bmatrix},$$

i.e., each row $\mathbf{c}_i^T$ is a right shifted version of the first row $\mathbf{c}_0^T$ of the matrix $\mathbf{C}$.
By definition, eigenvector satisfies the following equality

$$\mathbf{y} = \mathbf{C}\mathbf{w}_k = \lambda_k \mathbf{w}_k$$

The product $\mathbf{C}\mathbf{w}$ results in a column-vector $\mathbf{y}$ of length N, where the entry y_i, corresponds to the inner product of the vector $\mathbf{w}$ and the row $\mathbf{c}_i^T$.
Since

$$\mathbf{w}_k = \left(1, e^{-j\frac{2\pi}{N}k}, ..., e^{-j\frac{2\pi}{N}(N-1)k}\right)^T$$

and the row $\mathbf{c}_i^T$ of the matrix $\mathbf{C}$ is

$$(c_{N-i}, c_{N-i+1}, ..., c_{N-1}, c_0, c_1, ..., c_{N-i-1})^T,$$

the element y_i of $\mathbf{y}$ is then

$$y_i \;=\; \mathbf{c}_i^T \mathbf{w}_k = \sum_{m=N-i}^{N-1} c_m e^{-j\frac{2\pi}{N}(m-(N-i))k} + \sum_{m=0}^{N-i-1} c_m e^{-j\frac{2\pi}{N}(m+i)k}$$

Now, knowing that the complex exponential term e^{-jw} is periodic with period 2π, we can write the following

$$e^{-j\frac{2\pi}{N}mk} = e^{-j\frac{2\pi}{N}(m+N)k}$$

and the exponential terms of the first sum then reads $e^{-j\frac{2\pi}{N}(m-(N-i))k} = e^{-j\frac{2\pi}{N}(m+i)k}$. Finally, the inner product y_i is given by

$$y_i \;=\; \sum_{m=0}^{N-1} c_m e^{-j\frac{2\pi}{N}(m+i)k} = e^{-j\frac{2\pi}{N}ik} \sum_{m=0}^{N-1} c_m e^{-j\frac{2\pi}{N}mk} = e^{-j\frac{2\pi}{N}ik} C[k]$$

where $C[k]$ is the k-th Fourier coefficient of the sequence $\mathbf{c}_0$. Therefore, the product $\mathbf{y}$ is given by:

$$\mathbf{y} \;=\; \mathbf{C}\mathbf{w}_k = C[k]\left(1, e^{-j\frac{2\pi}{N}k}, ..., e^{-j\frac{2\pi}{N}(N-1)k}\right) = C[k]\mathbf{w}_k = \lambda_k \mathbf{w}_k$$

The Fourier basis vector $\mathbf{w}_k$ is, therefore, eigenvector with the corresponding Fourier coefficient $C[k]$ as its eigenvalue.

1.3 Basics of discrete signal processing

Exercise 8. Discrete sinc function
A discrete sinc function is the inverse DFT (IDFT) of the indicator function $I_M[n]$ of an interval $[-M, M]$, that is:

$$I_M[n] = \begin{cases} 1 & -M \leq n \leq M \\ 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

Assume that $2M$ is a divisor of N, with N the length of the IDFT.

(a) Derive a formula for $\text{sinc}_M[n]$.

(b) Using Matlab, compute the discrete sinc function and compare it with the result from (a).

Solution 8.

Exercise 9. Shannon and orthonormal basis
Shannon's sampling theorem states: A real bandlimited signal $f(t)$ having no spectral components equal or above ω_m is uniquely defined by its samples taken at twice ω_m, often called the Nyquist frequency. By denoting $T_s = \pi/\omega_m$, a reconstruction formula that complements the sampling theorem is:

$$f(t) = \sum_{n=-\infty}^{+\infty} f(nT_s)\mathrm{sinc}_{T_s}(t - nT_s), \tag{1.2}$$

where

$$\mathrm{sinc}_{T_s}(t) = \frac{\sin(\pi t/T_s)}{\pi t/T_s}. \tag{1.3}$$

An alternative interpretation of the sampling theorem is as a series expansion of bandlimited signals on an orthonormal basis. Define:

$$\varphi_{n,T_s}(t) = \frac{1}{\sqrt{T_s}}\mathrm{sinc}_{T_s}(t - nT_s). \tag{1.4}$$

(a) Show that $\{\varphi_{n,T_s}(t)\}_{n\in Z}$ form an orthonormal set, i.e.

$$< \varphi_{n,T_s}, \varphi_{m,T_s} >= \delta_{nm}. \tag{1.5}$$

(b) Show that any continuous-time signal $f(t)$ bandlimited to ω_m can be represented in the orthonormal basis $\{\varphi_{n,T_s}(t)\}_{n\in Z}$. That is, another way to write the interpolation formula (1.2) is

$$f(t) = \sum_{n=-\infty}^{+\infty} < \varphi_{n,T_s}, f >^* \varphi_{n,T_s}(t).$$

Solution 9. Shannon and orthonormal basis
(a) To prove that $\langle \varphi_n, \varphi_m \rangle = \delta_{nm}$, we use the Parseval's relation,

$$\begin{aligned}\langle \varphi_n, \varphi_m \rangle &= \int_{-\infty}^{+\infty} \varphi_n(t)\, \varphi_m^*(t)\, dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Psi_n(\omega)\, \Psi_m^*(\omega)\, d\omega\end{aligned}$$

where $\Psi(\omega)$ is the Fourier transform of $\varphi(t)$. We also know that the sinc function is equivalent to a rectangular function in the Fourier domain,

$$F\{\mathrm{sinc}_{T_s}(t)\} = \begin{cases} T_s & , \quad -\frac{\pi}{T_s} < \omega < \frac{\pi}{T_s} \\ 0 & , \quad \text{otherwise}\end{cases}$$

and therefore,

$$\Psi_n(\omega) = F\left\{\frac{1}{\sqrt{T_s}}\mathrm{sinc}_{T_s}(t - nT_s)\right\} = \begin{cases} \sqrt{T_s}e^{-j\omega n T_s} & , \quad -\frac{\pi}{T_s} < \omega < \frac{\pi}{T_s} \\ 0 & , \quad \text{otherwise}\end{cases}$$

Using this result in the Parseval's relation, we get

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{+\infty}\Psi_n(\omega)\,\Psi_m^*(\omega)\,d\omega &= \frac{1}{2\pi}\int_{-\frac{\pi}{T_s}}^{+\frac{\pi}{T_s}}\sqrt{T_s}e^{-j\omega nT_s}\sqrt{T_s}e^{+j\omega mT_s}d\omega \\
&= \frac{T_s}{2\pi}\int_{-\frac{\pi}{T_s}}^{+\frac{\pi}{T_s}} e^{j\omega(m-n)T_s}d\omega \\
&= \frac{T_s}{2\pi}\cdot\frac{1}{j(m-n)T_s}\left[e^{j\omega(m-n)T_s}\right]_{-\frac{\pi}{T_s}}^{+\frac{\pi}{T_s}} \\
&= \frac{1}{2\pi j(m-n)}\left[e^{j\pi(m-n)}-e^{-j\pi(m-n)}\right] \\
&= \frac{\sin[\pi(m-n)]}{\pi(m-n)} \\
&= \operatorname{sinc}(m-n)=0\ ,\ \ m\neq n
\end{aligned}$$

For $n=m$, we get

$$\begin{aligned}
\frac{T_s}{2\pi}\int_{-\frac{\pi}{T_s}}^{+\frac{\pi}{T_s}} e^{j\omega(m-n)T_s}d\omega &= \frac{T_s}{2\pi}\int_{-\frac{\pi}{T_s}}^{+\frac{\pi}{T_s}} 1d\omega \\
&= \frac{T_s}{2\pi}\left[\omega\right]_{-\frac{\pi}{T_s}}^{+\frac{\pi}{T_s}} \\
&= \frac{T_s}{2\pi}\left[\frac{\pi}{T_s}+\frac{\pi}{T_s}\right] \\
&= 1
\end{aligned}$$

Therefore,

$$\langle\varphi_n,\varphi_m\rangle=\tfrac{1}{2\pi}\textstyle\int_{-\infty}^{+\infty}\Psi_n(\omega)\,\Psi_m^*(\omega)\,d\omega=\delta_{nm}$$

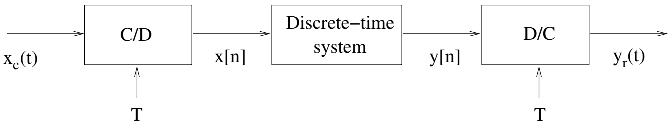
as stated.

(b) We wish to prove that $\frac{1}{\sqrt{T_s}}\langle\varphi_n,f\rangle^*=f(nT_s)$. We can easily do this by using the Parseval's relation, such that

$$\begin{aligned}
\frac{1}{\sqrt{T_s}}\langle\varphi_n,f\rangle^* &= \left(\frac{1}{2\pi\sqrt{T_s}}\int_{-\frac{\pi}{T_s}}^{+\frac{\pi}{T_s}}\sqrt{T_s}e^{-j\omega nT_s}F^*(\omega)\,d\omega\right)^* \\
&= \frac{1}{2\pi}\int_{-\omega_m}^{+\omega_m}F(\omega)\,e^{j\omega nT_s}d\omega \\
&= f(nT_s)
\end{aligned}$$

by definition of the Fourier transform.

Exercise 10. DISCRETE TIME PROCESSING
Consider the following system:



where the discrete-time system is a squarer, i.e., $y[n] = x^2[n]$.
What is the largest value of T such that $y_r(t) = x_c^2(t)$? Assume that $x_c(t)$ has the maximal frequency f_{max}.

Solution 10. DISCRETE TIME PROCESSING
Using the convolution-multiplication property, we have

$$\begin{aligned} y[n] &= x^2[n] \\ Y(e^{jw}) &= X(e^{jw}) \star X(e^{jw}) \end{aligned}$$

Therefore, $Y(e^{jw})$ will occupy twice the frequency band that $X(e^{jw})$ does if no aliasing occurs. Hence, $Y(e^{jw}) \neq 0$ for $-\pi < w < \pi$, implies $X(e^{jw}) \neq 0$ for $-\frac{\pi}{2} < w < \frac{\pi}{2}$. If we denote by f_{max} the maximum frequency of $X(e^{jw})$, then

$$\frac{\pi}{2} \geq 2\pi T f_{max}$$

and

$$T \leq \frac{1}{4 f_{max}}$$

Exercise 11. DOWNSAMPLING
Consider $x[n]$ and $y[n] = x[nN]$ as two sampled versions of the same continuous-time signal, with sampling periods T and NT, respectively. Prove that

$$Y(e^{j\omega}) = \frac{1}{N}\sum_{k=0}^{N-1} X(e^{j(\omega-2\pi k)/N}) \tag{1.6}$$

by going back to the underlying time-domain signal and resampling it with an N-times longer sampling period.
Hint: Recall that the discrete-time Fourier transform $X(e^{j\omega})$ of $x[n]$ is:

$$X(e^{j\omega}) \;=\; X_T\left(\frac{\omega}{T}\right) \;=\; \frac{1}{T}\sum_{k=-\infty}^{\infty} X_C\left(\frac{\omega}{T} - k\frac{2\pi}{T}\right), \tag{1.7}$$

where T is the sampling period. Then $Y(e^{j\omega}) = X_{NT}(\omega/NT)$ (since the sampling period is now NT), where $X_{NT}(\omega/NT)$ can be written similarly to (1.7). Finally, split the sum involved in $X_{NT}(\omega/NT)$ into $k = nN + l$, and gathering terms, (1.6) will follow.

Solution 11. DOWNSAMPLING
Consider $x[n]$ and $y[n]$ to be obtained from sampling $x_c(t)$ with sampling periods T and NT, respectively.

$$\begin{aligned}
Y(e^{jw}) &= X_{NT}\left(\frac{w}{NT}\right) = \frac{1}{NT}\sum_k X_c\left(\frac{w}{NT} - k\frac{2\pi}{NT}\right) \\
&= \frac{1}{NT}\sum_{l=0}^{N-1}\sum_{n=-\infty}^{\infty} X_c\left(\frac{w}{NT} - (nN+l)\frac{2\pi}{NT}\right) \; by\, defining\, n\, and\, l\, through\, k = nN + l \\
&= \frac{1}{N}\sum_{l=0}^{N-1}\left[\frac{1}{T}\sum_{n=-\infty}^{\infty} X_c\left(\frac{w-2\pi l}{NT} - n\frac{2\pi}{T}\right)\right] \\
&= \frac{1}{N}\sum_{l=0}^{N-1} X(e^{j(w-2\pi l)/N})
\end{aligned}$$

Exercise 12. Sampling & Quantization
We consider that each one of the two music instruments plays a tone. Consequently, the microphone will output a signal $x(t)$ composed of the sum of two tones, one at frequency f_1 and the other at frequency f_2, that is, $x(t) = A\cos(2\pi f_1 t) + B\cos(2\pi f_2 t)$.
As a numeric example we will consider $f_1 = 440.00$ Hz (a A4 note) and $f_2 = 523.25$ Hz (a C5 note), and $A = B = 1$.
This is just an example, and the acquisition system does not know *a priori* these numerical values.
The output of the microphone is sampled using a sampling frequency $f_s = 1$ kHz, obtaining a digital signal $x[n]$.

a) Sketch the magnitude of the spectrum of the analog signal $x(t)$.

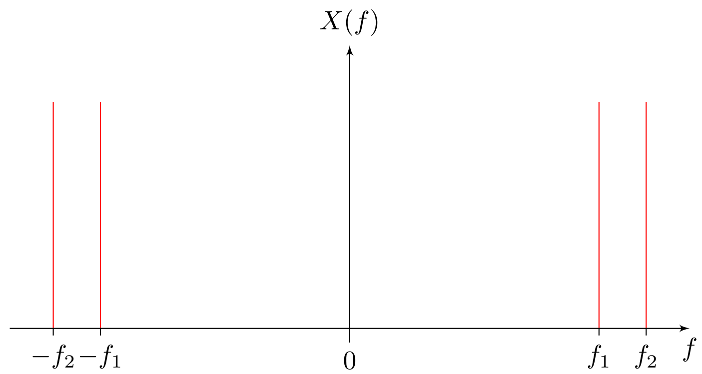
b) Sketch the magnitude of the spectrum of the digital signal $x[n]$ and clearly indicate the period (of length f_s) of such spectrum (choose the period that is symmetric with respect to the origin).

c) Using the sampling frequency $f_s = 1$ kHz, is the signal correctly sampled? (that is, is the information of the signal preserved?). Precisely justify your answer.

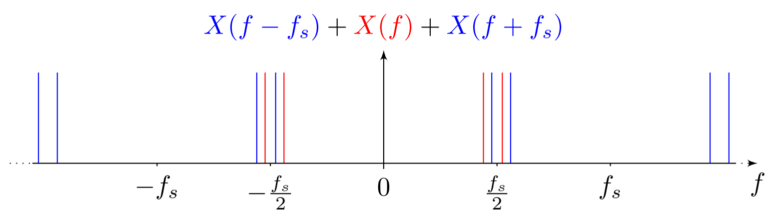
d) Given the digital signal $x[n]$ (sampled with $f_s = 1$ kHz), we reconstruct the analog signal using a $[-\frac{f_s}{2}, \frac{f_s}{2}]$ low pass filter. What kind of signal do we obtain? Which are the frequencies of the two tones of the reconstructed signal? Precisely justify your answer. Call f_a and f_b such frequencies.

Solution 12.

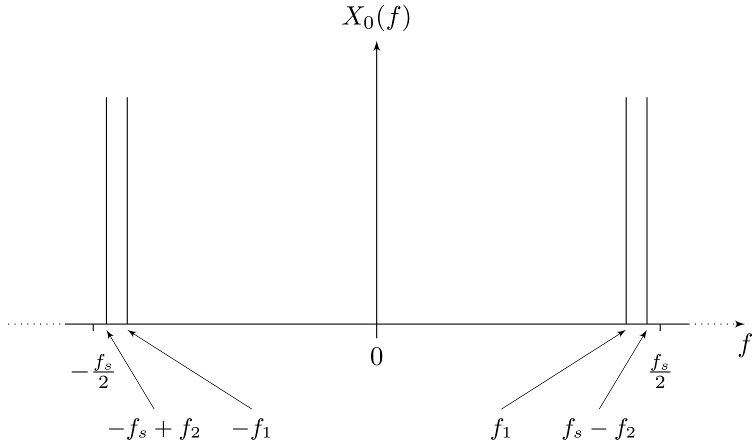
a) Magnitude of the spectrum of the analog signal:



b) Magnitude of the spectrum of the digital signal $x[n]$ (one replica)



The period is



c) In order to preserve the information of the analog signal, the latter must be sampled with a sampling frequency greater (or equal) than two times its maximum frequency $f_s \geq 2 \times f_{\max}$.

Here we have $f_s = 1$ kHz $< 2 \times f_{\max} = 2 \times 523.25$ Hz. Consequently the information of the analog signal is not preserved. It is indeed the case since the spectrum of the analog signal differs form the period of the spectrum of the sampled signal (see previous questions).

d) A $[-\frac{f_s}{2}, \frac{f_s}{2}]$ low pass filtering will provide an analog signal with a spectrum given by the period of the spectrum of the sampled signal. The so obtained analog signal corresponds to two sinusoids with frequencies $f_a = f_1 = 440.00$ Hz and $f_b = f_s - f_2 = 1000 - 523.25 = 476.75$ Hz, *i.e.*, $\widetilde{x}(t) = A\cos(2\pi f_a t) + B\cos(2\pi f_b t)$.

e) The signal amplitude goes from a minimum value of -1 to a maximum value of 1 (therefore the absolute maximum values is 1). The amplitude interval to be quantized is [-1,1], that is, an interval of length 2.

The 265 target values are the center of intervals of length $\Delta = 2/256$. The maximum absolute value of the quantization error is $\Delta/2 = 1/256$.

1.4 Probability

Exercise 13. Links between definitions

In this exercise we try to tackle what definitions imply other definitions. For each statement below, show if it is always true or not. (If it is false, a counter-example is sufficient.)

(a) The Power Spectral Density of a real-valued process is also real-valued.

(b) If a stochastic process is SSS, then the random variables in the process are i.i.d.

(c) If two random variables are independent, they are uncorrelated.

Solution 13. Links between definitions

(a) **True** If a WSS process is real valued, then it's (auto)correlation is symmetric. Since the Power Spectral Density is defined as the (Discrete Time) Fourier Transform of the correlation of the process:

$$S_X(\omega) = \sum_{k=-\infty}^{\infty} R_X[k]e^{-ik\omega},$$

we can use the property that the Fourier Transform of a symmetric signal is real valued.

Alternatively, we can use the intuition from the lecture, that the Power Spectral Density is the expected value of the square of the Fourier Transform of the signal, and thus it's real valued.

Additional remark, not in the scope of the class: if you want to define the Power Spectral Density for a non stationary signal, you have to do this locally, because the signal is going to change over time, and you can't rely on R_X. You can define the Power Spectral Density as exactly the expected value of the square of the Fourier Transform of the signal on some interval:

$$\begin{aligned}\hat{S}_{X,N}(\omega) &= \mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} X[k]e^{-jk\omega}\right|^2 \\ &= \frac{1}{N}\mathbb{E}\left(\sum_{k=0}^{N-1}\sum_{m=0}^{N-1} X[k]X^*[m]e^{-j\omega(k-m)}\right)\end{aligned}$$

Then, if you assume that the signal is actually stationary, you can simplify this expression as follows:

$$\begin{aligned}\hat{S}_{X,N}(\omega) &= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}\left(X[k]X^*[m]e^{-j\omega(k-m)}\right) \\ &= \sum_{l=-N+1}^{N-1} \frac{N-l}{N} R_X[l]e^{-jl\omega} \xrightarrow[N\to\infty]{} \sum_{l=-\infty}^{\infty} R_X[l]e^{-jl\omega}\end{aligned}$$

And therefore in the limit we get our "standard" PSD:

$$\hat{S}_{X,N}(\omega) \xrightarrow[N\to\infty]{} S_X(\omega)$$

which formalises intuition that the Power Spectral Density is exactly the expected value of the square of the Fourier Transform.

(b) **False**. A SSS process implies that it's values are indeed identically distributed, but *not always* independent. A simple counterexample is the discrete process built as follows:

$$X[n] = X[0] = Y \text{ for every } n$$

where Y is a (non constant) random variable. The process is clearly SSS, the variables Y are identically distributed, but are dependent.

(c) **True** It follows from the properties of expected value. If X and Y are *independent*, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, and therefore:

$$\mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)\big) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0,$$

so X and Y are uncorrelated.

Exercise 14. GAUSSIAN RANDOM VARIABLE
Suppose that a measurement $X[n]$ is affected by a random noise $W_1[n]$ due to external interferences and by a random noise $W_2[n]$ due to a defected measurement device. The noise $W_1[n]$ is i.i.d., distributed as a Gaussian random variable with mean m_1 and variance σ_1^2, and the noise $W_2[n]$ is i.i.d., distributed as a Gaussian random variable with mean m_2 and variance σ_2^2. Call W the sum of the two noises, *i.e.*, $W = W_1 + W_2$.

(a) Compute the mean and the variance of W;

(b) Give the joint distribution of W.

Suppose now that we don't know the law of the process $W[n]$ but we have observed a realization of it, $w[1], \dots, w[K]$. We are interested in estimating its mean based on the observation $w[1], \dots, w[K]$.

(c) Propose an empirical estimator of the mean;

(d) Check if such an estimator is biased.

Solution 14. GAUSSIAN RANDOM VARIABLE
(a) We compute the mean and the variance as:

$$E[W] = E[W_1 + W_2] = E[W_1] + E[W_2] = m_1 + m_2 = m$$

$$var(W) = E\left[|W_1 + W_2|^2\right] - |E\left[W_1 + W_2\right]|^2$$

For simplicity, we assume that W_1 and W_2 are real. Then,

$$\begin{aligned} var(W) &= E\left[|W_1|^2\right] + 2E\left[W_1 W_2^*\right] + E\left[|W_2|^2\right] - |E\left[W_1\right] + E\left[W_2\right]|^2 \\ &= E\left[|W_1|^2\right] + 2E\left[W_1\right]E\left[W_2^*\right] + E\left[|W_2|^2\right] - |E\left[W_1\right]|^2 - |E\left[W_2\right]|^2 - 2\left|E\left[W_1\right]E\left[W_2\right]\right| \\ &= var(W_1) + var(W_2) \end{aligned}$$

(b) The process $W[n]$ is a Gaussian process with mean $m = m_1 + m_2$ and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

(c) The natural way to estimate the mean is to compute the average of the realizations $w[k]$,

$$\hat{m}\left(w[1], ..., w[K]\right) = \frac{1}{K}\sum_{k=1}^{K} w[k]$$

(d) We analyze the bias

$$E\left[\hat{m}\left(w[1], ..., w[K]\right)\right] - m$$

$$lim_{N\to\infty}\left(E\left[\tfrac{1}{K}\sum_{k=1}^{K} w[k]\right] - m\right) = lim_{N\to\infty}E\left[\tfrac{1}{K}\sum_{k=1}^{K} w[k]\right] - m = 0$$

The bias is zero, and we say that the estimator is unbiased.

Exercise 15. SUM OF POISSON RV'S
Consider a sequence of iid random variables X_i, $i = 1, \ldots, N$, where each random variable has a Poisson distribution with mean λ. Consider now the random variable $Y = \sum_{i=1}^{N} X_i$.

(a) Compute the characteristic function $\mathbb{E}[e^{itX}]$ of a random variable X with a Poisson distribution with mean λ.

(b) Find the distribution of the random variable Y (Hint: You can use the characteristic function found in (a)).

(c) Find the mean and the variance of the random variable Y.

Solution 15.

(a) The characteristic function of a Poisson random variable X can be determined in the following way:

$$\begin{aligned}\mathbb{E}[e^{itX}] &= \sum_{k=0}^{\infty} P(X=k)e^{itk} = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} e^{itk} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} \\ &= e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)} \end{aligned} \tag{1.8}$$

(b) Taking $Y = \sum_{k=1}^{N} X_i$, and considering the fact that random variables X_i, $i = 1, \ldots, N$ are iid with Poisson distribution with mean λ, we have

$$\mathbb{E}[e^{itY}] = \mathbb{E}[e^{it \sum_{k=1}^{N} X_k}] \stackrel{iid}{=} \prod_{k=1}^{N} \mathbb{E}[e^{itX_k}] = \prod_{k=1}^{N} e^{\lambda(e^{it}-1)} = e^{N\lambda(e^{it}-1)} . \tag{1.9}$$

Since the characteristic function completely defines the probability distribution of a random variable, by comparing the final expressions in (1) and (2), we can see that the random variable Y has a Poisson distribution with mean $\lambda_Y = N\lambda$.

(c) The mean and the variance of the random variable Y are equal to λ_Y, where $\lambda_Y = N\lambda$.

Exercise 16. NEUROBIOLOGICAL SIGNAL AS A POISSON PROCESS

Consider the measurement of neurobiological spikes. We model the neurobiological spikes as a random process wherein the inter-arrival time between two spikes is an exponentially distributed random variable and is furthermore completely independent of other inter-arrival times. More precisely, assuming the spikes occur at $T_1, T_2, \ldots, T_K$, the time interval $T_i - T_{i-1} = t$, for $i = 1, 2, \ldots, K$, (assume that $T_0 = 0$) has the probability density function given by

$$p(T_i - T_{i-1} = t) = \lambda e^{-\lambda t}$$

for some $\lambda > 0$ and is zero for $t < 0$.

(a) Find the distribution of the number of spikes observed in the time interval $[0, t]$. More precisely, if $N(t)$ is the number of spikes in the time interval $[0, t]$, find $P\{N(t) = k\}$.

(b) What is the mean and the variance of the number of spikes in the interval $[0, t]$?

Furthermore, we define a Poisson process with the rate λ as follows:

The number of events/arrivals $N(t)$ in a finite interval of time t obeys the Poisson distribution

$$P\{N(t) = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t} .$$

Moreover, the number of arrivals $N(t_1, t_2)$ in the time interval $[t_1, t_2]$ is independent of the number of arrivals $N(t_3, t_4)$ in the time interval $[t_3, t_4]$, if the time intervals do not overlap.

Assume now that we know that the signal of neurobiological spikes can be described as a Poisson process with the rate λ.

(c) What is distribution of spikes' inter-arrival times? Are they independent of each other?

Solution 16.

(a) We want to compute $P\{N(t) = k\}$, i.e. the probability that exactly k spikes occur in the interval $[0, t]$. Let the first spike occur at instant t_1, the second at t_2 and so on (the k^{th} spike occurs at instant t_k). Let $s_1, s_2, \ldots, s_k, s_{k+1}$ denote the corresponding inter-arrival times, which are independent by the problem assumption. Note that s_{k+1} denotes the inter-arrival time between the $(k+1)^{th}$ and k^{th} spike, and that the $(k+1)^{th}$ spike should occur outside the interval $[0, t]$. Putting it all together, we have:

$$
\begin{aligned}
&P\{N(t) = k\} \\
&= P\{t_1 \le t,\, t_1 \le t_2 \le t,\, \ldots,\, t_{k-1} \le t_k \le t,\, t_{k+1} > t\} \\
&= P\{s_1 \le t,\, s_2 \le t - s_1,\, \ldots,\, s_k \le t - \sum_{l=1}^{k-1} s_l,\, s_{k+1} > t - \sum_{l=1}^{k} s_l\} \\
&= P\{s_1 \le t\}\, P\{s_2 \le t - s_1\} \ldots P\{s_k \le t - \sum_{l=1}^{k-1} s_l\}\, P\{s_{k+1} > t - \sum_{l=1}^{k} s_l\} \\
&= \int_0^t p(s_1) \int_0^{t-s_1} p(s_2) \ldots \int_0^{t-\sum_{l=1}^{k-1} s_l} p(s_k) \int_{t-\sum_{l=1}^{k} s_l}^{\infty} p(s_{k+1})\, \mathrm{d}s_{k+1} \mathrm{d}s_k \ldots \mathrm{d}s_2 \mathrm{d}s_1 \\
&= \int_0^t \lambda e^{-\lambda s_1} \int_0^{t-s_1} \lambda e^{-\lambda s_2} \ldots \int_0^{t-\sum_{l=1}^{k-1} s_l} \lambda e^{-\lambda s_k} \int_{t-\sum_{l=1}^{k} s_l}^{\infty} \lambda e^{-\lambda s_{k+1}}\, \mathrm{d}s_{k+1} \mathrm{d}s_k \ldots \mathrm{d}s_2 \mathrm{d}s_1 \\
&= \lambda^{k+1} \int_0^t e^{-\lambda s_1} \int_0^{t-s_1} e^{-\lambda s_2} \ldots \int_0^{t-\sum_{l=1}^{k-1} s_l} e^{-\lambda s_k} \left(-\frac{1}{\lambda} e^{-\lambda s_{k+1}} \right) \Bigg|_{t-\sum_{l=1}^{k} s_l}^{\infty} \mathrm{d}s_k \ldots \mathrm{d}s_2 \mathrm{d}s_1 \\
&= \lambda^{k+1} \int_0^t e^{-\lambda s_1} \int_0^{t-s_1} e^{-\lambda s_2} \ldots \int_0^{t-\sum_{l=1}^{k-1} s_l} e^{-\lambda s_k} \left(\frac{1}{\lambda} e^{-\lambda (t - \sum_{l=1}^{k} s_l)} \right) \mathrm{d}s_k \ldots \mathrm{d}s_2 \mathrm{d}s_1 \\
&= \lambda^k e^{-\lambda t} \int_0^t \int_0^{t-s_1} \ldots \int_0^{t-\sum_{l=1}^{k-1} s_l} \mathrm{d}s_k \ldots \mathrm{d}s_2 \mathrm{d}s_1 \\
&= \frac{\lambda^k t^k}{k!} e^{-\lambda t}
\end{aligned} \tag{1.10}
$$

We can see from (1.10) that the number of spikes in the interval $[0, t]$ has a Poisson distribution with mean λt.

(b) The mean and the variance are both equal to λt.

(c) Call by S the random variable denoting the time interval between two consecutive arrivals. The event $\{S > t\}$ is equivalent to the event $\{N(t) = 0\}$. Thus, $P\{S > t\} = P\{N(t) = 0\} = e^{-\lambda t}$, which implies that $P\{S \le t\} = 1 - e^{-\lambda t}$. In other words, S is an exponentially distributed random variable. Furthermore, the inter-arrival times are independent

$$P\{T_{k+1} - T_k > t | T_k - T_{k-1} = s\} = P\{T_{k+1} - T_k > t\}\,,$$

since the number of arrivals in the time interval $(T_{k+1}, T_k]$ is independent of the number of arrivals in the time interval $(T_k, T_{k-1}]$ (the interval $(T_k, T_{k+1}]$ does not overlap with the time interval $(T_{k-1}, T_k]$).

Exercise 17. POISSON PROCESS
Consider two neurons, A and B. Action of a stimulus to neuron A (respectively, B) generates a spike train which can be modeled as a Poisson process A_t (respectively, B_t) with rate λ_A (respectively, λ_B).

1) Assume that the neurons do not interact and behave independently on the action of a stimulus. Additionally, suppose we stimulate both neurons at the same time. What is the statistics of the resultant spike train?

2) Assume that neuron B is a complete clone of neuron A. The two neurons fire simultaneously, but we suppose to be still able to count two overlapping spikes. What is the statistic of the resultant spike train? (Hint: Is the process $2A_t$ a Poisson process?)

Solution 17.

1) Denote by $A(t)$ and $B(t)$ the number of spikes in the time interval of length t. Since the two neurons behave independently, we have

$$\begin{aligned} P(A(t)+B(t)=k) &= \sum_{l=0}^{k} P(A(t)=l)\, P(B(t)=k-l) \\ &= \sum_{l=0}^{k} \frac{e^{-t\lambda_A}(t\lambda_A)^l}{l!} \frac{e^{-t\lambda_B}(t\lambda_B)^{k-l}}{(k-l)!} \\ &= \sum_{l=0}^{k} e^{-t\lambda_A}(t\lambda_A)^l\, e^{-t\lambda_B}(t\lambda_B)^{k-l} \binom{k}{l} \frac{1}{k!} \\ &= \frac{e^{-t\lambda_A}e^{-t\lambda_B}}{k!} \sum_{l=0}^{k} \binom{k}{l} (t\lambda_A)^l (t\lambda_B)^{k-l} \\ &= \frac{e^{-t(\lambda_A+\lambda_B)}}{k!} (t(\lambda_A+\lambda_B))^k\,. \end{aligned}$$

Thus, the resultant spike train is a Poisson process with mean $\lambda_A+\lambda_B$.

2) Since the two neurons are identical, the spike train has the distribution of the random variable $2A$. Clearly, $2A$ is not a Poisson process. Indeed, $P(2A(t)=2k+1)=P(A(t)=k+1/2)=0$, since A takes only non-negative integer values. Thus, $2A$ has support of only even numbers. Assuming $k=2n$, we get

$$P(2A(t)=k)=P(A(t)=n)=\frac{e^{-t\lambda_A}(t\lambda_A)^n}{n!}=\frac{e^{-t\lambda^A}(t\lambda_A)^{k/2}}{(k/2)!}\,.$$

Check that

$$\sum_{k=0}^{\infty} P(2A(t)=k)=\sum_{n=0}^{\infty} P(2A(t)=2n)=\sum_{n=0}^{\infty}\frac{e^{-t\lambda_A}(t\lambda_A)^n}{n!}=1\,.$$

Exercise 18.
What is the correlation $R_X[n]$ of an independent identically distributed (i.i.d) process X of variance $\sigma_X^2 = 1$ and zero mean? What is the power spectral density $S_X(\omega)$?

Solution 18.
The correlation is $[R_X[n] = m_X^2 + \sigma_X^2 \delta[n] = \delta[n]$. The power spectral density is $S_X(\omega) = \mathcal{F}[R_X[n]] = 1$.

Exercise 19. Power Spectrum Density
Consider the stochastic process defined as

$$Y[n] = X[n] + \beta X[n-1]$$

where $\beta \in \mathbb{R}$ and $X[n]$ is a zero-mean wide-sense stationary process with autocorrelation function given by

$$R_X[k] = \sigma^2 \alpha^{|k|}$$

for $|\alpha| < 1$.

(a) Compute the power spectrum density $S_Y(e^{j\omega})$ of $Y[n]$.

(b) For which values of β does $Y[n]$ corresponds to a white noise? Explain.

Solution 19. Power Spectrum Density
The process $Y[n]$ is obtained by filtering the wide-sense stationary process $X[n]$, i.e.

$$Y[n] = H(z)X[n],$$

with $H(z) = 1 + \beta z^{-1}$. Therefore,

$$S_Y(\omega) = |H(e^{j\omega})|^2 S_X(\omega).$$

The function $|H(e^{j\omega})|^2$ is given by

$$|H(e^{j\omega})|^2 = 1 + 2\beta\cos\omega + \beta^2.$$

The PSD of $X[n]$ is computed by taking the DTFT of $R_X[n]$, that gives

$$S_X(\omega) = \sum_{k=-\infty}^{\infty} R_X[k]e^{-j\omega k} = \sigma_X^2 \frac{1-\alpha^2}{1-2\alpha\cos\omega+\alpha^2}.$$

Hence, the PSD of $Y[n]$ is

$$S_Y(\omega) = \sigma_X^2(1-\alpha^2)\frac{1+2\beta\cos\omega+\beta^2}{1-2\alpha\cos\omega+\alpha^2}.$$

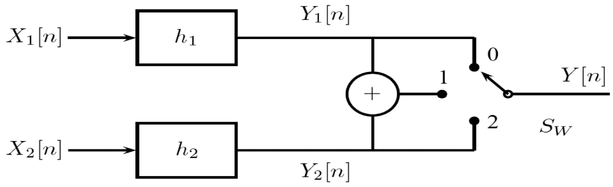
To have that $Y[n]$ is a white process, we should impose that the spectral density is a constant. This corresponds to setting $\beta = -\alpha$. The interpretation is the following. The process $X[n]$ is an AR process. In fact, it can be obtained by filtering a white noise $W[n]$, which has variance $\sigma_X^2(1-\alpha^2)$, with the synthesis filter

$$H_s(z) = \frac{1}{1-\alpha z^{-1}},$$

which has a pole for $z = \alpha$. The filter $H(z)$ is an FIR filter (i.e. it has only zeros) and has exactly one zero at $z = -\beta$. We can imagine that the process $Y[n]$ is obtained by filtering $W[n]$ with the cascade of the filters $H_s(z)$ and $H(z)$. Therefore, to obtain a white noise at the output, we must have that the zero of $H(z)$ cancels the pole of $H_s(z)$, i.e. $-\beta = \alpha$.

Exercise 20. Stationarity

Consider the following block diagram



The two input processes $X_1[n]$, $X_2[n]$ are jointly gaussian, uncorrelated, white, with zero mean, and variance $\sigma^2_{X_1} = 1$, $\sigma^2_{X_2} = 2$ respectively. The two blocks h_1, h_2 are linear, time-invariant filters with transfer functions

$$H_1(z) = 1 + z^{-1},$$
$$H_2(z) = 1 - z^{-1},$$

respectively. The output process $Y[n]$ is obtained by means of the output switch S_W.

(a) Suppose that the output switch is constantly on the position "0," or "2," what can you say on the output process $Y[n]$? Is it stationary in wide and/or strict sense? Is it gaussian? Compute the correlation and (if it exists) the spectral density of $Y[n]$.

(b) What happens if the switch is on the position "1"? Answer to the same question of the previous case.

(c) Suppose that the position of the switch changes with the value of the time index n. The switch takes the position "0," when n is even and the position "1," when n is odd. Is $Y[n]$ stationary in this case? Is it gaussian? Compute the correlation function of $Y[n]$ (be careful on the definition of the correlation function in this case!)

(d) Suppose that the switch takes a random position among "0" and "1" with equal probability and independently of the values of $X_1[n]$ and $X_2[n]$. Is $Y[n]$ stationary in some sense in this case? What is the correlation and (if it exists) the spectral density of $Y[n]$?

Solution 20. Stationarity

(a) The autocorrelation of $h_1[n]$ is given by

$$\begin{aligned}
r_{h_1}[m] &= h_1[m] \star h_1[-m] \\
&= (\delta[m] + \delta[m-1]) \star (\delta[-m] + \delta[-m-1]) \\
&= (\delta[m] + \delta[m-1]) \star (\delta[m] + \delta[m+1]) \\
&= \delta[m] \star \delta[m] + \delta[m] \star \delta[m+1] \\
&\quad + \delta[m-1] \star \delta[m] + \delta[m-1] \star \delta[m+1] \\
&= \delta[m-1] + 2\delta[m] + \delta[m+1]
\end{aligned}$$

Since $X[n]$ is a white process, $r_{x_1}[m] = \sigma_x^2 \delta[m]$, and thus

$$\begin{aligned}
r_{y_1}[m] &= r_{h_1}[m] \star r_{x_1}[m] \\
&= (\delta[m-1] + 2\delta[m] + \delta[m+1]) \star \sigma_x^2 \delta[m] \\
&= \sigma_x^2 \delta[m-1] + 2\sigma_x^2 \delta[m] + \sigma_x^2 \delta[m+1] \\
&= \delta[m-1] + 2\delta[m] + \delta[m+1]
\end{aligned}$$

The power spectral density $S_{y_1}(\omega)$ is the Fourier transform of the autocorrelation function $r_{y_1}[m]$,

$$\begin{aligned} S_{y_1}(\omega) &= F\{r_{y_1}[m]\} \\ &= e^{-j\omega} + 2 + e^{j\omega} \\ &= 2 + 2\cos\omega \end{aligned}$$

Another way of obtaining $S_{y_1}(\omega)$ is by performing all calculations in the z domain:

$$\begin{aligned} S_{y_1}(z) &= S_{h_1}(z)\, S_{x_1}(z) \\ &= \left[H_1(z)\, H_1\left(z^{-1}\right)\right] \sigma_x^2 \\ &= (1+z^{-1})(1+z) \\ &= z^{-1} + 2 + z \end{aligned}$$

and then setting $z = e^{j\omega}$ for obtaining the Fourier transform:

$$\begin{aligned} S_{y_1}(\omega) &= e^{-j\omega} + 2 + e^{j\omega} \\ &= 2 + 2\cos\omega \end{aligned}$$

For the second case, we get

$$r_{y_2}[m] = -2\delta[m-1] + 4\delta[m] - 2\delta[m+1]$$

$$S_{y_2}(\omega) = 4 - 4\cos\omega$$

(b) Since X_1 and X_2 are uncorrelated, Y_1 and Y_2 are also uncorrelated, and thus

$$\begin{aligned} r_y[m] &= \mathbb{E}[Y[n]\, Y[n+m]] \\ &= \mathbb{E}[(Y_1[n] + Y_2[n])\,(Y_1[n+m] + Y_2[n+m])] \\ &= \mathbb{E}[Y_1[n]\, Y_1[n+m]] + \mathbb{E}[Y_2[n]\, Y_2[n+m]] \\ &\quad + \mathbb{E}[Y_1[n]\, Y_2[n+m]] + \mathbb{E}[Y_2[n]\, Y_1[n+m]] \\ &= \mathbb{E}[Y_1[n]\, Y_1[n+m]] + \mathbb{E}[Y_2[n]\, Y_2[n+m]] + 0 + 0 \\ &= r_{y_1}[m] + r_{y_2}[m] \\ &= -\delta[m-1] + 6\delta[m] - \delta[m+1] \end{aligned}$$

And the power spectral density is given by

$$\begin{aligned} S_y(\omega) &= F\{r_y[m]\} \\ &= -e^{-j\omega} + 6 - e^{j\omega} \\ &= 6 - 2\cos\omega \end{aligned}$$

which is equivalent to $S_{y_1}(\omega) + S_{y_2}(\omega)$.
(c) We have four possible cases:

$$\mathbb{E}\left[y\left[k\right]y\left[l\right]\right] = \begin{cases} \mathbb{E}\left[y_1\left[k\right]y_1\left[l\right]\right] & , \quad k,l \text{ even} \\ \mathbb{E}\left[\left(y_1\left[k\right]+y_2\left[k\right]\right)\left(y_1\left[l\right]+y_2\left[l\right]\right)\right] & , \quad k,l \text{ odd} \\ \mathbb{E}\left[y_1\left[k\right]\left(y_1\left[l\right]+y_2\left[l\right]\right)\right] & , \quad k \text{ even, } l \text{ odd} \\ \mathbb{E}\left[\left(y_1\left[k\right]+y_2\left[k\right]\right)y_1\left[l\right]\right] & , \quad k \text{ odd, } l \text{ even} \end{cases}$$

which, according to the results above, equals

$$\mathbb{E}\left[y\left[k\right]y\left[l\right]\right] = \begin{cases} r_{y_1}\left[k-l\right] & , \quad k,l \text{ even} \\ r_{y_1}\left[k-l\right]+r_{y_2}\left[k-l\right] & , \quad k,l \text{ odd} \\ r_{y_1}\left[k-l\right] & , \quad k \text{ even, } l \text{ odd} \\ r_{y_1}\left[k-l\right] & , \quad k \text{ odd, } l \text{ even} \end{cases}$$

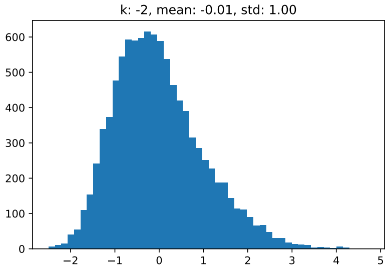
which means that the process is not stationary anymore.
(d) In this case, the autocorrelation function $r_y\left[k-l\right]$ takes into account the random variable that determines the position of the switch. Since the probability is equally distributed, each combination of k and l occurs with a probability of $\frac{1}{4}$. Hence,

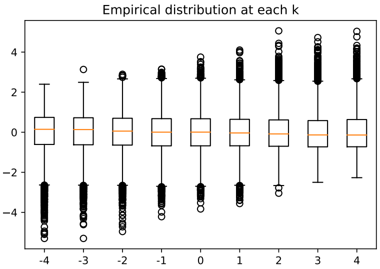
$$\begin{aligned} r_y\left[k-l\right] &= \frac{3}{4}r_{y_1}\left[k-l\right]+\frac{1}{4}\left(r_{y_1}\left[k-l\right]+r_{y_2}\left[k-l\right]\right) \\ &= r_{y_1}\left[k-l\right]+\frac{1}{4}r_{y_2}\left[k-l\right] \end{aligned}$$

which means that the process is, in fact, stationary.

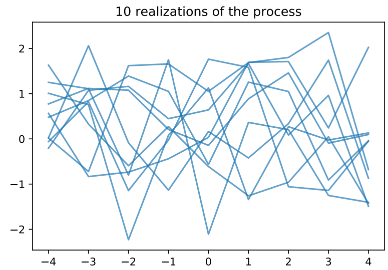
Exercise 21. Coding: Wide Sense Stationary Processes vs. Stationary Processes
The goal of this exercise is to assess the stationarity of a random process via common tools from the field of descriptive statistics.
Follow the instructions in `Stationarity.ipynb`.

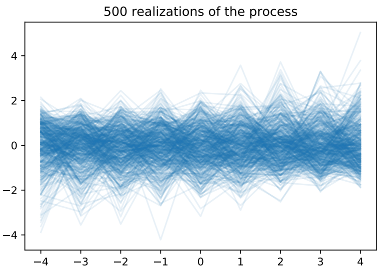
Solution 21. Coding: Wide Sense Stationary Processes vs. Stationary Processes
For example code see Jupyter notebook, available on Moodle.





By looking at an example histogram of (normalised) skewed distribution (left), we can see that it's not symmetric. This will change with k.
Another way to look at the distributions is boxplot (right). It's sometimes more convenient than histogram. The orange horizontal lines in the middle are the medians, the boxes depicts the "middle" 50% of the distributions, the vertical lines cover most of the distribution and the dots depict "unlikely" points, outliers. You can see that those distributions change in time, and that the median and the main mass, that is middle 50% of the distribution does not move that much, but the placement of the outliers what matters. If you look closely you can see that the median moves slightly in the opposite direction that the outliers do.





If you look at just a few (for example 10) realizations of this process it's hard to judge if it a stationary (gaussian) noise, or if it is some more complicated distriburion. If you look at 500 realisations you can see some pattern emerge that distinguish the "past" and the "future" (negative and positive k).

Exercise 22. CORRELATION AND DEPENDENCE
Give an example of a two random variables X and Y that are uncorrelated (their covariance is equal to 0) but they are dependent.

(a) Write down the distribution of pair (X, Y) and marginal distributions of X and Y.

(b) [Coding] Sample the joint distribution of (X, Y) and plot those samples. Then sample from X and Y independently and plot resulting samples. Do those plots look similar?

(b) Explain your example to your neighbour(s).

(c) Check your neighbour's example. Are their variables dependent?

Hint: it might be easier to come up with discrete distributions.

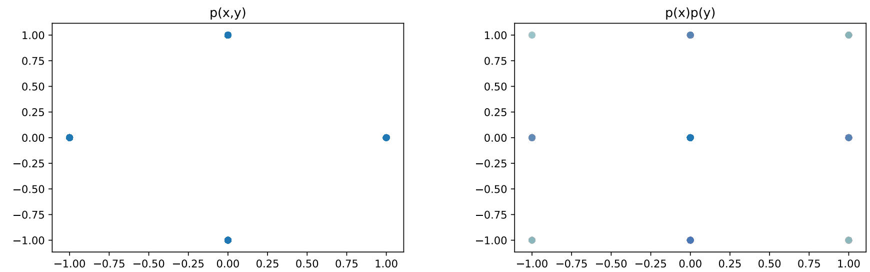
Solution 22.
Let us consider a very simple pair of two discrete variables, each taking values in $\{-1, 0, 1\}$. We want X to be zero if and only if Y is not zero. This way they are obviously dependent, but we have $\mathbb{E}(XY) = 0$. We only need now the product $\mathbb{E}(X)\mathbb{E}(Y) = 0$. For that we just need the probabilities of $X = 1$ and $X = -1$ to be equal.
Based on that we can construct many different pairs, but one of them would be $p(x, y)$:

		x		
		-1	0	1
y	-1	0	0.25	0
	0	0.25	0	0.25
	1	0	0.25	0

Then the marginal probabilities are the same for X and Y:

$$p(x) = \begin{cases} 0.25 & \text{for } x = -1 \\ 0.5 & \text{for } x = 0 \\ 0.25 & \text{for } x = 1 \end{cases}$$

We can then plot the joint and product distributions side-by-side and assess their resemblance.



Another example with Gaussian RVs is described on Wikipedia.

Exercise 23. MIXTURE MODEL
Consider an i.i.d. signal $X[n]$ which takes only two possible values, m_1 and m_2. When we try to measure such a signal we obtain only a noisy version of it

$$Y[n] = X[n] + W[n]\,,$$

where $W[n]$ is a centered Gaussian white noise with power σ_W^2, independent of X.

(a) Show that Y is wide sense stationary.

(b) Give the distribution of $Y[n]$ (PDF or CDF).

- Hint 1: you can use the law of total probability, which in this case means that:

$$F(Y[n]) = F(Y[n], X[n] = m_1) + F(Y[n], X[n] = m_2)$$

- Hint 2: you can use Bayes' rule, which in this case means that:

$$F(Y[n], X[n]) = F(Y[n]|X[n])P(X[n])$$

(c) Download data file `MixtureModel.csv` from Moodle. It was generated from this process using some unknown m_2, m_1 and σ_W. The probability of $X[n] = m_1$ was $\frac{1}{2}$. Load the data in Matlab or Python, and take a look. Why is this called "two class mixture model"?

(d) Guess the parameters m_i and propose a method that based on the value of $Y[n]$ guesses a value of $X[n]$ (no need for a formal description).

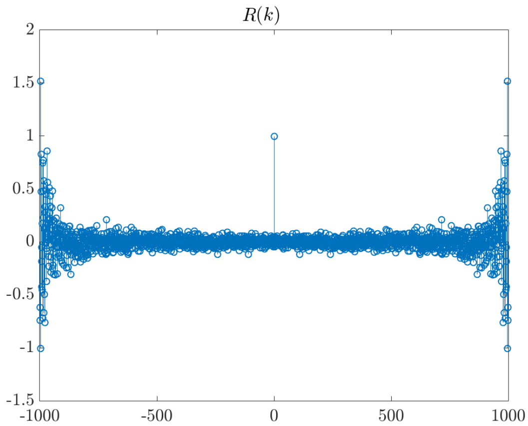
(e) Based on your guess of m_i calculate σ_W.

Solution 23.

Exercise 24. Correlation (4pts)
Let $W[n]$ be a centered white noise with $\sigma^2 = 1$, taking real values. Given that the process is i.i.d. and centered, we know that, theoretically,

$$R(k) = \mathbb{E}\left[W[k+n]W[n]\right] = \begin{cases} \sigma^2 = 1 & k = 0\,; \\ 0 & k \neq 0\,; \end{cases}$$

We have measured $N = 1000$ samples of the noise $w[1], \ldots, w[1000]$ and then we have computed the correlation $R(k)$, $k = -999, \ldots, 0, \ldots, 999$. Here's the plot of the correlation.



Can you tell if the plotted correlation has been computed using:
The empirical un-biased correlation

$$R(k) = \frac{1}{N - |k|} \sum_{n=1}^{N-|k|} w[n+k]w[n]\,, \quad k = -(N-1), \ldots, (N-1)\,,$$

or the empirical biased correlation

$$R(k) = \frac{1}{N} \sum_{n=1}^{N-|k|} w[n+k]w[n]\,, \quad k = -(N-1), \ldots, (N-1)\,.$$

Precisely justify your answer.

Solution 24. Correlation
The plotted correlation has been computed using the empirical un-biased correlation:

$$R(k) = \frac{1}{N - |k|} \sum_{n=1}^{N-|k|} w[n+k]w[n]\,, \quad k = -(N-1), \ldots, (N-1)\,.$$

We can tell it by the high variance for k close to 1000, that appears due to smaller number of measurements k samples apart. The biased correlation would have dumped this variance by basing the tails towards zero (bias-variance trade-off).

Chapter 2

ARMA Models

Exercise 25. Projection of AR process
Consider the autoregressive process defined as:

$$X[n] = X[n-1] - X[n-2] + 2X[n-3] + W[n], \tag{2.1}$$

where $W[n]$ is white noise. Your task is to find the best *linear* predictor $Y[n-1]$ of order 3 of $X[n+2]$ given the past $(X[n-1], X[n-2], X[n-3])$, using two methods:

(a) by using the projection theorem and solving a system of linear equations,

(b) by recursively expanding the definition of $X[n+2]$ and using an intuitive property.

Solution 25.
We will find a predictor of $X[n+2]$ using two different methods and see that they yield the same result. First, from the definition of X, we know that:

$$X[n+2] = X[n+1] - X[n] + 2X[n-1] + W[n+2] \tag{2.2}$$

(a) Projection theorem.

We want to find $Y[n-1]$ such that $\mathbb{E}\big((X[n+2] - Y[n-1])Y[n-k]\big)$ for $k = 1, 2, 3$. We can write Y as:

$$Y[n-1] = \sum_{k=1}^{3} a_k X[n-k]$$

To find a projection, we need to solve the system of equations (for $k = 1, 2, 3$), so that the difference between our predictor Y and process X is orthogonal "to the past":

$$\mathbb{E}\big((X[n+2] - a_1X[n-1] - a_2X[n-2] - a_3X[n-3])X[n-k]\big) = 0$$

In order to calculate actual values of a_i, we have to first rewrite the system in terms of the auto-correlation:

$$R_X[2+k] - a_1R_X[k-1] - a_2R_X[k-2] - a_3R_X[k-3] = 0. \tag{2.3}$$

On the other hand, for an AR process we have Yule-Walker equations:

$$R_X[m] - \sum_k p_k R_X[m-k] = \delta_m S_W^2$$

Where in our case $p_1 = 1,\ p_2 = -1$ and $p_3 = 2$.

From these equations we find $R_X[m]$ for $m = 0, 1, \ldots 5$. Those values can then be plugged into (2.3). Finally, (2.3) can be solved for a_i.

(b) Intuitive property.

We again start with (2.2). We need to expand terms dependent on n and $n+1$ and write them in terms of the past. We first expand $X[n+1]$:

$$\begin{aligned} X[n+2] &= X[n+1] - X[n] + 2X[n-1] + W[n+2] \\ &= (X[n] - X[n-1] + 2X[n-2] + W[n+1]) - X[n] + 2X[n-1] + W[n+2] \\ &= X[n-1] + 2X[n-2] + W[n+1] + W[n+2]. \end{aligned}$$

The first two terms of the equations are functions of the past, and the last two terms are innovations that are entirely independent of the past. Therefore, using intuitive property, we get that Y, the best linear predictor of X is:

$$Y[n-1] = X[n-1] + 2X[n-2]$$

We can see that the second solution is much easier on paper.

Exercise 26. A Simple AR Process
Consider the discrete time stochastic process $\{X[n]\}_{n\geq 0}$ defined by

$$X[n+1] = aX[n] + W[n+1], \quad n \geq 0$$

where $|a| < 1$, $X[0]$ is a Gaussian random variable of mean 0 and variance c^2, and $\{W[n]\}_{n\geq 1}$ is a sequence of i.i.d. Gaussian variables of mean 0 and variance σ^2, and independent of $X[0]$.

(a) Express $X[n]$ in terms of $X[0], W[1], \ldots, W[n]$ (and a). Give the mean and variance of $X[n]$;

(b) Suppose now that $c^2 = \frac{\sigma^2}{1-|a|^2}$. Show that with this specific choice for the variance of $X[0]$, the process $\{X[n]\}_{n\geq 0}$ is strictly stationary.

(c) Give the one-step predictor of $X[n]$: $\widehat{X}[n|n-1]$.

(d) What is the whitening (or analysis) filter of $\{X[n]\}_{n\in\mathbb{Z}}$?
What is the generating (or synthesis) filter of $\{X[n]\}_{n\in\mathbb{Z}}$?

(e) Give the covariance function $R_X[k] = \mathbb{E}\left[X[n+k]X[n]^*\right]$.

(f) Write $X[n]$ in terms of $W[n]$, $W[n-1]$ and $X[n-2]$. Deduce from this the two-step predictor of $X[n]$: $\widehat{X}[n|n-2]$, the projection of $X[n]$ onto $H(X, n-2)$, the Hilbert subspace spanned by the random variables $X[n-2], X[n-3], \ldots$.

Solution 26. A Simple AR Process

(a) The recursion formula

$$X[n+1] = aX[n] + W[n+1], \quad n \geq 0$$

yields

$$\begin{aligned}
X[1] =& aX[0] + W[1] \\
X[2] =& aX[1] + W[2] = a^2X[0] + aW[1] + W[2] \\
X[3] =& aX[2] + W[3] = a^3X[0] + a^2W[1] + aW[2] + W[3] \\
&\vdots \\
X[n] =& a^nX[0] + a^{n-1}W[1] + a^{n-2}W[2] + \cdots + W[n] \\
=& a^nX[0] + \sum_{k=0}^{n-1} a^kW[n-k].
\end{aligned}$$

Hence the mean of $X[n]$ is given by

$$\mathbb{E}\left[X[n]\right] = a^n\mathbb{E}\left[X[0]\right] + \sum_{k=0}^{n-1} a^k\mathbb{E}\left[W[n-k]\right] = 0$$

since both $\mathbb{E}\left[X[0]\right] = 0$ and $\mathbb{E}\left[W[j]\right] = 0, \quad \forall j \geq 0$.

The variance of $X[n]$ is given by

$$\begin{aligned}\mathbb{E}\left[|X[n]|^2\right] &= \mathbb{E}\left[\left(a^n X[0] + \sum_{k=0}^{n-1} a^k W[n-k]\right)\left(a^n X[0] + \sum_{j=0}^{n-1} a^j W[n-j]\right)^*\right] \\ &= |a|^{2n}\mathbb{E}\left[|X[0]|^2\right] + \sum_{j=0}^{n-1} a^n a^{*j}\mathbb{E}\left[X[0]W^*[n-j]\right] \\ &+ \sum_{k=0}^{n-1} a^k a^{*n}\mathbb{E}\left[W[n-k]X^*[0]\right] \\ &+ \sum_{k,j=0}^{n-1} a^k a^{*j}\mathbb{E}\left[W[n-k]W^*[n-j]\right]\end{aligned}$$

Recall that $\mathbb{E}\left[|X[0]|^2\right] = c^2$, the sequence of random variables $W[n]$ and $X[0]$ are independent and centered, thus $\mathbb{E}\left[X[0]W^*[n-j]\right] = \mathbb{E}\left[W[n-k]X^*[0]\right] = 0, \quad \forall 0 \le j, k \le (n-1)$ and $\mathbb{E}\left[W[n-k]W^*[n-j]\right] = \sigma^2\delta[j-k]$. Combining these observations, we have

$$\begin{aligned}\mathbb{E}\left[|X[n]|^2\right] &= |a|^{2n}c^2 + \sigma^2\sum_{k,j=0}^{n-1} a^k a^{*j}\delta[j-k] \\ &= |a|^{2n}c^2 + \sigma^2\sum_{k=0}^{n-1}|a|^{2k} \\ &= |a|^{2n}c^2 + \sigma^2\left(\frac{1-|a|^{2n}}{1-|a|^2}\right)\end{aligned}$$

(b) If $c^2 = \frac{\sigma^2}{1-|a|^2}$, the variance of $X[n]$ is independent of n and is given by

$$\mathbb{E}\left[|X[n]|^2\right] = \frac{\sigma^2}{1-|a|^2}.$$

Following the same steps in part (a), one can show that

$$X[n+k] = a^k X[n] + a^{k-1}W[n+1] + a^{k-2}W[n+2] + \cdots + W[k+n]. \tag{2.4}$$

Thus

$$\begin{pmatrix} X[n] \\ X[n+1] \\ \vdots \\ X[n+k] \end{pmatrix} = A \begin{pmatrix} X[n] \\ W[n+1] \\ \vdots \\ W[n+k] \end{pmatrix}$$

The distribution of $(X[n], W[n+1], \cdots, W[n+k])$ is independent of $n \ge 0$, and therefore the distribution of $(X[n], \cdots, X[n+k])$ is independent of $n \ge 0$, therefore $\{X[n]\}_{n\ge 0}$ is strictly stationary.

(c) Recall that

$$X[n] - aX[n-1] = W[n],$$

thus $\langle X[n] - aX[n-1], u\rangle = \langle W[n], u\rangle = 0$ for all $u \in H(X, n-1)$ since $H(X, n-1) = H(W, n-1)$ (see Theorem 2.2 in class notes). Recall that, roughly speaking, $H(W, n-1)$

is composed of linear combinations of $W[n-1], W[n-2], \ldots$. Note also that $aX[n-1] \in H(X, n-1)$ (since it is a linear function of $X[n-1]$), hence by the projection theorem this is the best linear approximation (best in least square sense) for $X[n]$, thus $\widehat{X}[n|n-1] = aX[n-1]$.

(d) The whitening filter makes $\{X[n]\}_{n\geq 0}$ a white noise, here we have

$$X[n] - aX[n-1] = W[n],$$

so it is clear that $P(z) = 1 - az^{-1}$. The generating filter is given by

$$H^s(z) = \frac{1}{P(z)} = \frac{1}{(1-az^{-1})} = \sum_{n\geq 0} a^n z^{-n}$$

and

$$X[n] = W[n] + aW[n-1] + a^2 W[n-2] + \cdots + a^k W[n-k] + \ldots$$

(e) Using (2.4) we obtain

$$\begin{aligned} \mathbb{E}\left[X[n+k]X^*[n]\right] =& \mathbb{E}\left[\left(a^k X[n] + \sum_{j=0}^{k-1} a^j W[n+k-j]\right) X^*[n]\right] \\ =& a^k \mathbb{E}\left[|X[n]|^2\right] + \sum_{j=0}^{k-1} a^j \mathbb{E}\left[W[n+k-j]X^*[n]\right] \\ =& a^k \mathbb{E}\left[|X[n]|^2\right] \\ =& a^k \left(|a|^{2n} c^2 + \sigma^2 \left(\frac{1-|a|^{2n}}{1-|a|^2}\right)\right) \end{aligned}$$

where the last equality follows from part (a) and $\mathbb{E}\left[W[n+k-j]X^*[n]\right] = 0$ since $W[n+k-j]$ and $X[n]$ are independent $\forall 0 \leq j \leq k-1$. Plugging $c^2 = \frac{\sigma^2}{1-|a|^2}$ yields

$$R_X[k] = a^k \frac{\sigma^2}{1-|a|^2}.$$

Note that the above equality together with the fact that $\mathbb{E}\left[X[n]\right] = 0$ shows that the process $\{X[n]\}_{n\geq 0}$ is wide sense stationary with the special condition $c^2 = \frac{\sigma^2}{1-|a|^2}$. Since the process is wide sense stationary and Gaussian, it is strictly stationary. Recall that the statistics of a Gaussian process is completely determined by its first and second order properties.

(f) Again from (2.4), we have

$$X[n] = a^2 X[n-2] + aW[n-1] + W[n].$$

thus $\langle X[n] - a^2 X[n-2], u\rangle = \langle aW[n-1], u\rangle + \langle W[n], u\rangle = a\langle W[n-1], u\rangle + \langle W[n], u\rangle = 0$ for all $u \in H(X, n-2)$ since $H(X, n-2) = H(W, n-2)$ and the random variables $W[n-2], W[n-3], \ldots$ are independent of $W[n-1]$ and $W[n]$. Note also that $a^2 X[n-2] \in H(X, n-2)$, hence by the projection theorem this is the best least square approximation for $X[n]$ knowing $X[n-2], X[n-3], \ldots$, thus $\widehat{X}[n|n-2] = a^2 X[n-2]$.

Exercise 27. CANONICAL REPRESENTATION
Let $\{X[n]\}_{n\in\mathbb{Z}}$ be a centered AR signal with power spectral density

$$S_X(\omega) = \frac{1}{26 - 10\cos\omega}$$

(a) Give the *canonical* representation of it

$$P(z)\, X[n] = W[n]$$

(give $P(z)$, give the variance σ^2 of $\{W[n]\}_{n\in\mathbb{Z}}$).
Hint: you can use Féjer's identity, i.e., for all $\beta \in \mathbb{C}$, $\beta \neq 0$, and for all $z \in \mathbb{C}$ such that $|z| = 1$:

$$(z-\beta)\left(z - \frac{1}{\beta^*}\right) = -\frac{1}{\beta^*}\, z|z-\beta|^2.$$

(b) Give the one-step predictor of $X[n]$: $\widehat{X}[n|n-1]$.

(c) What is the whitening (or analysis) filter of $\{X[n]\}_{n\in\mathbb{Z}}$?
What is the generating (or synthesis) filter of $\{X[n]\}_{n\in\mathbb{Z}}$?

(d) Give the covariance function $R_X[k] = \mathbb{E}\left[X[n+k]X[n]^*\right]$.

(e) Write $X[n]$ in terms of $W[n]$, $W[n-1]$ and $X[n-2]$. Deduce from this the two-step predictor of $X[n]$: $\widehat{X}[n|n-2]$, the projection of $X[n]$ onto $H(X, n-2)$, the Hilbert subspace spanned by the random variables $X[n-2], X[n-3], \ldots$.

Solution 27. CANONICAL REPRESENTATION

(a)

$$\begin{aligned} S_X(\omega) &= \frac{1}{26 - 10\cos\omega} = \frac{1}{26 - 10\frac{e^{j\omega}+e^{-j\omega}}{2}} \\ &= \frac{1}{26 - 5e^{j\omega} - 5e^{-j\omega}} = \frac{1}{(5e^{j\omega}-1)(5e^{-j\omega}-1)} \\ &= \frac{1}{5e^{j\omega}(1-\frac{1}{5}e^{-j\omega})5e^{-j\omega}(1-\frac{1}{5}e^{j\omega})} = \frac{1}{25(1-\frac{1}{5}e^{-j\omega})(1-\frac{1}{5}e^{j\omega})} \end{aligned}$$

Recalling that

$$S_X(\omega) = \frac{1}{|P(e^{j\omega})|^2}\sigma^2$$

one can choose

$$P(z) = 1 - \frac{1}{5}z^{-1}$$

and $\sigma^2 = \frac{1}{25}$. Note that with this choice $P(z)$ is minimum phase, that is, it is stable, causal and all its zeros lie inside the unit circle.

(b) With above choice for $P(z)$, we have the following recursion for $X[n]$,

$$X[n] - \frac{1}{5}X[n-1] = W[n] \tag{2.5}$$

where $\{W[n]\}_{n\in\mathbb{Z}}$ is white noise sequence with variance $\sigma^2 = \frac{1}{25}$. The exact reasoning in the Solution of Exercise 3, part (b) will yield $\widehat{X}[n|n-1] = \frac{1}{5}X[n-1]$.

(c) The whitening filter is

$$P(z) = 1 - \frac{1}{5}z^{-1}$$

and the generating filter is

$$H^s(z) = \frac{1}{P(z)} = \frac{1}{(1 - \frac{1}{5}z^{-1})} = \sum_{n \geq 0} \left(\frac{1}{5}\right)^n z^{-n}.$$

(d) Using the above expression for the generating filter yields

$$X[n] = W[n] + \frac{1}{5}W[n-1] + \cdots + \left(\frac{1}{5}\right)^k W[n-k] + \ldots$$

and

$$X[n+k] = W[n+k] + \cdots + \left(\frac{1}{5}\right)^k W[n] + \left(\frac{1}{5}\right)^{(k+1)} W[n-1] + \ldots$$

Since $\mathbb{E}\left[W[i]W[j]^*\right] = \sigma^2 \delta_{ij}$, the only non-zero terms in $\mathbb{E}\left[X[n+k]X[n]^*\right]$ are those with $i = j$. Therefore,

$$R_X[k] = \mathbb{E}\left[X[n+k]X[n]^*\right] = \left(\frac{1}{5}\right)^k \sigma^2 \left(1 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^4 + \ldots\right) = \left(\frac{1}{5}\right)^k \frac{1}{25}\frac{25}{24}.$$

(e) The recursion in (2.5) yields

$$X[n] = \frac{1}{5}X[n-1] + W[n] = \left(\frac{1}{5}\right)^2 X[n-2] + \frac{1}{5}W[n-1] + W[n].$$

$\widehat{X}[n|n-2] = \left(\frac{1}{5}\right)^2 X[n-2]$ since $\left(\frac{1}{5}\right)^2 X[n-2] \in H\left(X, n-2\right)$ (it is a linear function of $X[n-2]$) and $X[n] - \left(\frac{1}{5}\right)^2 X[n-2] = \frac{1}{5}W[n-1] + W[n] \perp H\left(X, n-2\right)$ (indeed, $\frac{1}{5}W[n-1] + W[n] \perp H\left(W, n-2\right)$ and $H\left(W, n-2\right) = H\left(X, n-2\right)$ for the canonical representation (see Theorem 2.2 in class notes)).

Exercise 28.

This exercise has a more theoretical flavor (no long computation).

We consider a voice synthesizer based on filtering of a white noise $H(z)W[n] = X[n]$, where the input $W[n]$ is a real Gaussian white noise, centered, with correlation $R_W[n] = \delta_n \alpha$ ($\alpha > 0$),

$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

is a minimum phase filter with real coefficients a_1 and a_2, and $X[n]$ is the process describing the synthesized voice.

(a) Is X[n] a wide-sense stationary process? Justify **precisely** your answer.

We are now interested in computing the second order properties of $X[n]$

(b) Compute the power spectral density of $X[n]$, $S_X(\omega)$.

(c) Compute the mean of $X[n]$ and express the variance of $X[n]$ as a function of $S_X(\omega)$.

(d) Give the recursive expression of the correlation of $X[n]$.

(e) Using the recursive expression of the correlation write a system of linear equations that allows to obtain the variance of $X[n]$ as a function of a_1, a_2 and α.

Hint: Exploit the fact that $X[n]$ is real.

In real life, the synthesizer is not perfect and its defects can be modelled as an additive white noise $V[n]$ with variance σ_V^2. Finally, the signal we obtain is given by

$$Y[n] = X[n] + V[n]$$

where $X[n]$ and $V[n]$ are supposed to be independent. In particular, if we call $\mathcal{H}(X)$ the Hilbert space spanned by $X[n]$ and $\mathcal{H}(V)$ the Hilbert space spanned by $V[n]$, the independence on X and V implies the orthogonality of the two spaces $\mathcal{H}(X)$ and $\mathcal{H}(V)$. Call now $\mathcal{H}(W)$ the Hilbert space spanned by $W[n]$

(f) Are the two spaces $\mathcal{H}(V)$ and $\mathcal{H}(W)$ orthogonal? Justify **precisely** your answer.

By listening the noised synthesized voice, we would like to estimate the characteristics of its generating system. More precisely:
A) From the process $Y[n]$ we would like to recover the original synthesized voice $X[n]$.

(g) Give the transfer function of a filter to optimally estimate, in the mean square sense, $X[n]$ from $Y[n]$, and express such a transfer function in terms of a_1, a_2, α, and σ_V^2.

B) From the estimate of $X[n]$ we would like to recover the coefficients a_1 and a_2 of the transfer function $H(z)$ and the parameter α of the white noise $W[n]$,

(h) Write the system of linear equations to obtain a_1, a_2, and α from the correlation of $X[n]$.

Solution 28.
The first part of the exercise can be straightforwardly solved by applying the fundamental filtering formula for wide-sense stationary processes. We recall that such formula requires the input process to be wide-sense stationary with summable correlation and the filter to be stable. The input process $W[n]$ is a white noise and it is very easy to check that is wide-sense stationary and with summable correlation. The filter is stable since, by assumption, is minimum phase, and it is time invariance, since by construction its coefficients are constant.
We remark that, in general, we cannot directly prove properties of $X[n]$ from the fact that $X[n]$ is a linear combination of $W[n]$. The key point is the stability of the linear combination, *i.e.*, the stability of the filter. If the filter is not stable or not time invariant, $X[n]$ is still a linear combination of $W[n]$, but $X[n]$ is surely not wide-sense stationary, its mean is surely not zero, and the Hilbert space its spans is surely not equal to the space spanned by $W[n]$.
We also remark that the exercise clearly asked for precise answers.

1) Yes, $X[n]$ is wide-sense stationary by the fundamental filtering formula for wide-sense stationary processes.

2) By the fundamental filtering formula

$$\begin{aligned} S_X(\omega) &= |H(\omega)|^2 S_W(\omega) \\ &= \frac{\alpha}{(1+a_1 e^{-j\omega}+a_2 e^{-2j\omega})(1+a_1 e^{j\omega}+a_2 e^{2j\omega})} \end{aligned} \tag{2.6}$$

Note that $S_W(\omega) = \alpha$ since $W[n]$ is a white noise.

3) Again, by the fundamental filtering formula

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[W\right] \sum_{k=-\infty}^{\infty} h[k].$$

$W[n]$ is centered, that is $\mathbb{E}[W] = 0$ and $\sum_{k=-\infty}^{\infty} h[k] < \infty$ since $H(z)$ is stable, thus $\mathbb{E}[X] = 0$.

Alternatively, one can study the specific structure of the problem. In symbolic notation, the filtering by $H(z)$ can be expressed as

$$\begin{aligned} X[n] &= H(z)W[n] \\ &= \frac{1}{1+a_1 z^{-1}+a_2 z^{-2}} W[n] \end{aligned}$$

which can be equivalently written as

$$W[n] = (1+a_1 z^{-1}+a_2 z^{-2})X[n].$$

Interpreting z^{-1} as the delay operator we have

$$W[n] = X[n] + a_1 X[n-1] + a_2 X[n-2] \tag{2.7}$$

which also reveals the auto-regressive structure of the process. Taking expectation on both sides of the above equality gives

$$\begin{aligned} \mathbb{E}\left[W[n]\right] &= \mathbb{E}\left[X[n]\right] + a_1 \mathbb{E}\left[X[n-1]\right] + a_2 \mathbb{E}\left[X[n-2]\right] \\ 0 &= \mathbb{E}\left[X\right] + a_1 \mathbb{E}\left[X\right] + a_2 \mathbb{E}\left[X\right] \end{aligned}$$

Thus $\mathbb{E}[X] = 0$. We remark that the right side of equation (2.7) is a finite linear combination of $X[n]$. Therefore, the expectation is just the sum of the expectations of each terms.

The filter $H(z)$ being stable and the correlation of the input process $R_W[n] = \alpha\delta[n]$ being summable, the correlation of the output process $R_X[n]$ is also summable. Hence, we have the following relation for the variance of $X[n]$

$$\mathrm{Var}(X) = R_X[0] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_X(\omega)d\omega.$$

The fact that $\mathrm{Var}(X) = R_X[0]$ is because of the fact that $X[n]$ is centered.

4) The correlation structure of $X[n]$ can be obtained by either using the formulas derived in class for AR processes, after observing that the process is an AR-process as we already did in part 3), or one can do the computation explicitly. Using the formula for AR processes will directly yield

$$R_X[m] + a_1 R_X[m-1] + a_2 R_X[m-2] = \alpha\delta[m], \qquad m \geq 0 \tag{2.8}$$

and $R_X[-m] = R_X[m]$ since $X[n]$ is real.

5) Considering the recursive expression in (2.8) for $m = 0, 1, 2$ together with the fact that $R_X[-m] = R_X[m]$ (since $X[n]$ is real) yields the following system of linear equations

$$\begin{aligned} R_X[0] + a_1 R_X[1] + a_2 R_X[2] &= \alpha \\ R_X[1] + a_1 R_X[0] + a_2 R_X[1] &= 0 \\ R_X[2] + a_1 R_X[1] + a_2 R_X[0] &= 0. \end{aligned} \tag{2.9}$$

The three equations can be easily solved to obtain the three unknowns $R_X[0]$, $R_X[1]$ and $R_X[2]$. By recalling that $\mathrm{Var}(X) = R_X[0]$, the solution of the system gives us the variance.

6) Due to the stability and time invariance of the filter, we know from class notes (see section "A Hilbert Space Viewpoint of Linear Prediction" of the ARMA chapter - in the current notes is theorem Theorem 2.2 of Chapter 2) that

$$\mathcal{H}(X) = \mathcal{H}(W)$$

(we remark that the result was presented for the Hilbert spaces spanned by the past, *i.e.*, $\mathcal{H}(X, n)$, but it is straightforwardly extended to the Hilbert spaced spanned by the whole processes). Since we know that $\mathcal{H}(X)$ and $\mathcal{H}(V)$ are orthogonal, $\mathcal{H}(W)$ and $\mathcal{H}(V)$ are also orthogonal.

Alternatively, from the fact that $W[n]$ is a finite linear combination of $X[n]$, $X[n-1]$ and $X[n-2]$ (see equation (2.7)) we have $\mathcal{H}(W) \subseteq \mathcal{H}(X)$. consequently, the orthogonality between $\mathcal{H}(X)$ and $\mathcal{H}(V)$ implies the orthogonality between $\mathcal{H}(W)$ and $\mathcal{H}(V)$.

7) The filter that optimally estimates $X[n]$ from $Y[n]$, optimally in the sense that it minimizes the mean squared error, is the Wiener filter whose transfer function is given by

$$F(e^{j\omega}) = \frac{S_{XY}(\omega)}{S_Y(\omega)}. \tag{2.10}$$

Recall that in our case,

$$Y[n] = X[n] + V[n]$$

and $X[n]$ and $V[n]$ are independent. The cross-correlation of $X[n]$ and $Y[n]$ is given by

$$\begin{aligned} R_{XY}[m] &= \mathbb{E}[X[n+m]Y[n]] \\ &= \mathbb{E}[X[n+m](X[n]+V[n])] \\ &= \mathbb{E}[X[n+m]X[n]] + \mathbb{E}[X[n+m]V[n]] \\ &= R_X[m] \end{aligned}$$

since $X[n]$ and $V[n]$ are independent. Taking the Fourier transform and using the expression for $S_X(\omega)$ given in (2.6), we obtain

$$\begin{aligned} S_{XY}(\omega) &= S_X(\omega) \\ &= \frac{\alpha}{(1+a_1e^{-j\omega}+a_2e^{-2j\omega})(1+a_1e^{j\omega}+a_2e^{2j\omega})}. \end{aligned} \tag{2.11}$$

The auto-correlation of $Y[n]$ is given by

$$\begin{aligned} R_Y[m] &= \mathbb{E}[Y[n+m]Y[n]] \\ &= \mathbb{E}[(X[n+m]+V[n+m])(X[n]+V[n])] \\ &= \mathbb{E}[X[n+m]X[n]] + \mathbb{E}[X[n+m]V[n]] + \mathbb{E}[V[n+m]X[n]] + \mathbb{E}[V[n+m]V[n]] \\ &= \mathbb{E}[X[n+m]X[n]] + \mathbb{E}[V[n+m]V[n]] \\ &= R_X[m] + R_V[m]. \end{aligned}$$

Taking the Fourrier transform, we obtain

$$\begin{aligned} S_Y(\omega) &= S_X(\omega) + S_V(\omega) \\ &= \frac{\alpha}{(1+a_1e^{-j\omega}+a_2e^{-2j\omega})(1+a_1e^{j\omega}+a_2e^{2j\omega})} + \sigma_V^2. \end{aligned} \tag{2.12}$$

Note that $S_V(\omega) = \sigma_V^2$ since $V[n]$ is white. Combining (2.10), (2.11) and (2.12) we obtain

$$F(e^{j\omega}) = \frac{\alpha}{\alpha + \sigma_V^2(1+a_1e^{-j\omega}+a_2e^{-2j\omega})(1+a_1e^{j\omega}+a_2e^{2j\omega})}.$$

8) Just observe that the system of linear equations in (2.9) can be solved for a_1, a_2 and α if the correlation of $X[n]$, that is if $R_X[0]$, $R_X[1]$ and $R_X[2]$ are known. Thus, the desired system of equations is the one given in (2.9). Note that we have referred this system of linear equations as Yule-Walker Equations in the class.

Chapter 3

Prediction and Estimation in the General Non-ARMA Case

Exercise 29.
The process $X[n]$ is a real AR process of order M.

1) Write the recursion that allows to synthesize the process $X[n]$ from a white noise process $W[n]$.
2) Write the correlation structure of X.

   Suppose that the order M is unknown but an estimate of the correlation function $R_X[n]$ is available for $n \geq 0$.

3) Describe precisely a procedure to determine the order M, the parameters of the AR model and the variance of the input noise W.
4) What is the expression of the power spectral density of $X[n]$, $S_X(\omega)$ (as a function of the parameters and the order M)?

Solution 29.

1) The recursion that allows to synthesize the process $X[n]$ from a white noise process $W[n]$ is:
$$X[n] = a_1 X[n-1] + \ldots + a_M X[n-M] + W[n].$$
2) The correlation structure of X is.
$$R_X[n] = a_1 R_X[n-1] + \ldots + a_M R_X[n-M] + \delta[n]\sigma_W^2.$$
3) According to Yule-Walker equations, compute the mean square error $\|\epsilon_m\|_2^2$ of the linear prediction for increasing orders m. This error will strictly decrease until order $M+1$, then it will be a constant. This is how you compute the order M of the ARMA process. Then you can solve the Yule-Walker equations of order$M$ to get the parameters of the AR model. The mean square prediction error of the predictor of order M corresponds to the variance of the input noise W.

4)
$$H(z) = \frac{1}{1 - a_1 z^{-1} - \ldots - a_M z^{-M}}.$$

Then $X(z) = H(z)W(z)$. Using the fundamental filtering formula, we obtain

$$S_X(\omega) = |H(e^{j\omega})|^2 \sigma_W^2.$$

Exercise 30.
The process $X[n]$ is a real AR process of order 2:

$$X[n] = \frac{1}{4}X[n-1] + \frac{1}{8}X[n-2] + W[n],$$

where W[n] is a white noise. Using Yule Walker equations, give the best linear predictor of order 1 of $X[n]$. Compare the obtained coefficient with 1/4 and comment.

Extra Training: Same exercise with $X[n] = 0.3X[n-1] + 0.7X[n-2] - 0.154X[n-3] + W[n]$, and a linear predictor of order 2. This exercise will make you handle the correlation matrix.

Solution 30.
We are looking for a such that

$$\hat{X}[n] = aX[n-1]$$

and that minimizes

$$\|\epsilon\|_2^2 = \mathbb{E}[|X[n] - \hat{X}[n]|^2].$$

According to Yule-Walker equations, $a = R_X[1]/R_X[0]$.

$$\begin{aligned} R_X[1] &= \mathbb{E}[X[n]X[n-1]] \\ &= \frac{1}{4}\mathbb{E}[X[n-1]X[n-1]] + \frac{1}{8}\mathbb{E}[X[n-2]X[n-1]] + \mathbb{E}[WX[n-1]] \\ &= \frac{1}{4}R_X[0] + \frac{1}{8}R_X[1]. \end{aligned}$$

We conclude that $a = 2/7$. $2/7$ is slightly larger than $1/4$. This can be explained by the positive covariance between $X[n-1]$ and $X[n-2]$. So this predictor says:

$$X[n] = \frac{1}{4}X[n-1] + \frac{1}{28}X[n-1].$$

$\frac{1}{28}X[n-1]$ predicts the contribution of $\frac{1}{8}X[n-2]$.

For the extra question,

$$\begin{aligned} a_1 &\simeq 0.196 \\ a_2 &\simeq 0.670. \end{aligned}$$

An important thing to know to solve this exercise is:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Exercise 31.
In this exercise, we consider the application of the Wiener filter in reducing additive noise. Consider a signal $X[n]$ embedded in additive zero mean white Gaussian noise. That is,

$$Y[n] = X[n] + W[n].$$

Assume that $X[n]$ and $W[n]$ are uncorrelated.

(a) Derive the transfer function of an optimal non-causal filter.

(b) We define the following signal to noise ratio:

$$\mathrm{a}(\omega) = \frac{R_{xx}(e^{j\omega})}{R_{ww}(e^{j\omega})}.$$

How is the Wiener filter response in the case of noise-free frequencies, i.e., $a(\omega_o) \gg 1$? and in the case of very high noise, i.e., $a(\omega_o) \approx 0$? what can you conclude?

Solution 31.

(a) The expression for the Wiener filter is

$$H(e^{j\omega}) = \frac{S_{XY}(\omega)}{S_Y(\omega)}$$

If $Y[n] = X[n] + W[n]$ and $X[n]$ and $W[n]$ are uncorrelated with $W[n]$ zero-mean, we have

$$R_{XY}[m] = \mathbb{E}\left[X[n+m](X[n] + W[n])\right] = \mathbb{E}\left[X[n+m]X[n]\right] = R_X[m]$$
$$R_Y[m] = \mathbb{E}\left[(X[n+m] + W[n+m])(X[n] + W[n])\right] = R_X[m] + R_W[m]$$

hence,

$$S_{XY}(\omega) = S_X(\omega)$$
$$S_Y(\omega) = S_X(\omega) + S_W(\omega).$$

The Wiener filter is thus

$$H(e^{j\omega}) = \frac{S_X(\omega)}{S_X(\omega) + S_W(\omega)}$$

(b)

$$H(e^{j\omega}) = \frac{S_X(\omega)}{S_X(\omega) + S_W(\omega)} = \frac{a(\omega)}{a(\omega) + 1}.$$

If $a(\omega_o) \gg 1$, $H(e^{j\omega}) \approx 1$. That is, the filter applies little or no attenuation to the noise-free frequency component. If $a(\omega_o) \approx 0$, $H(e^{j\omega}) \approx 0$. That is, the filter applies a high attenuation to the noisy frequency component. In conclusion, for additive noise, the Wiener filter attenuates each frequency component in proportion to an estimate of the signal to noise ratio.

Exercise 32.
Suppose that $X[n]$ is zero mean white Gaussian noise with variance $\sigma_X^2 = 1$. A desired response $D[n]$ is obtained by applying $X[n]$ to a linear filter $h[n]$; Our task is to design a linear filter $g[n]$ that minimizes a cost function J given by

$$J = \mathbb{E}\left[E^2[n]\right] = \mathbb{E}\left[(D[n] - Y[n])^2\right]. \tag{3.1}$$

Suppose that $h[n]$ is a 3-tap FIR filter given by $[h_0, h_1, h_2]$. We want to determine a 2-tap optimum Wiener filter, which minimizes the cost function J.

(a) Compute the cross correlation vector $R_{DX}[n]$.

(b) Determine an optimal 2-tap filter $g[n]$.

(c) Repeat (a) and (b) for the case when $h[n]$ is an IIR filter with a transfer function given by

$$H(z) = \frac{1}{1 - az^{-1}}. \tag{3.2}$$

Solution 32.

(a) The cross correlation vector $R_{DX}[m]$ is defined as:

$$\begin{aligned} R_{DX}[m] &= \mathbb{E}\left[D[n]X[n-m]\right] \\ &= \mathbb{E}\left[(h_0X[n] + h_1X[n-1] + h_2X[n-2])X[n-m]\right] \\ &= h_0R_X[m] + h_1R_X[m-1] + h_2R_X[m-2] \\ &= h_0\delta[m] + h_1\delta[m-1] + h_2\delta[m-2] \end{aligned}$$

where we have used the fact that $X[n]$ is a white noise with variance $\sigma_X^2 = 1$, Hence we find that

$$R_{DX}[m] = \begin{cases} h_0 & \text{if } m = 0 \\ h_1 & \text{if } m = 1 \\ h_2 & \text{if } m = 2 \\ 0 & \text{otherwise} \end{cases}$$

(b) The optimal filter that estimates $D[n]$ from $X[n]$ is given by the Wiener filter formula:

$$\sum_k g[k]R_X[m-k] = R_{DX}[m].$$

Recall from the derivation of the formula that the equation above is obtained by differantiating with respect to $g[m]$ (say the m'th tap of the wiener filter in the context of this exercise). Note that in this exercise we are restricted to use a two tap filter whose tap gains we would like to choose optimally. Thus, in order for the first tap gain $g[0]$ to be optimal, the filter should satisfy

$$\begin{aligned} \sum_k g[k]R_X[-k] &= R_{DX}[0] \\ g[0]\delta[0] + g[1]\delta[-1] &= h_0\delta[0] + h_1\delta[-1] + h_2\delta[-2] \end{aligned}$$

thus $g[0] = h_0$ and the optimality condition for the second tap gain yields

$$\sum_k g[k]R_X[1-k] = R_{DX}[1]$$
$$g[0]\delta[1] + g[1]\delta[0] = h_0\delta[1] + h_1\delta[0] + h_2\delta[-1]$$

hence $g[1] = h_1$.

(c) To calculate $R_{DX}[m]$ we need to use the formula

$$\frac{1}{1-az^{-1}} = 1 + az^{-1} + a^2z^{-2} + ... + a^nz^{-n} + ...$$

Therefore

$$D[n] = \sum_{k=0}^{\infty} a^k X[n-k]$$

The cross correlation vector $R_{DX}[m]$ is then

$$R_{DX}[m] = \mathbb{E}\left[\left(\sum_{k=0}^{\infty} a^k X[n-k]\right) X[n-m]\right]$$

$$= \sum_{k=0}^{\infty} a^k \delta[m-k]$$
$$= \begin{cases} a^m & \text{if } m \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

And by the same arguments in part (b) one can easily find $g[0] = 1$ and $g[1] = a$.

Exercise 33. WIENER FILTER
Suppose that a desired process $X[n]$ is generated by filtering the white gaussian noise $W[n]$ (centered with variance 1) using a filter $h[k]$, where

$$H(z) = \frac{1 + \frac{3}{4}z^{-1}}{1 + \frac{1}{2}z^{-1}}. \tag{3.3}$$

Consider now the signal $Y[n] = X[n] + V[n]$, where $V[n]$ is zero mean white Gaussian noise with variance 1/2 and uncorrelated with $W[n]$.

(a) Design a Wiener filter for estimating $X[n]$ from $Y[n]$.

(b) Repeat (a) for the case when $V[n]$ is a random variable given by:

$$V[n] = \frac{1}{3}X[n-1] - \frac{1}{9}X[n-2].$$

Solution 33.

(a) By filtering the unit-variance white noise with a filter $H(e^{j\omega})$, we get an output signal $X[n]$ with power spectrum:

$$S_X(\omega) = |H(e^{j\omega})|^2.$$

The Wiener filter that estimates $X[n]$ from $Y[n]$:

$$\begin{aligned} Q(e^{j\omega}) &= \frac{S_X(\omega)}{S_X(\omega) + S_V(\omega)} \\ &= \frac{|H(e^{j\omega})|^2}{|H(e^{j\omega})|^2 + (1/2)}. \end{aligned}$$

This is sufficient solution, but if we want to get exact formula, we can start with calculating the square norm of $H(e^{j\omega})$:

$$|H(e^{j\omega})|^2 = \frac{1 + \frac{3}{2}\cos(\omega) + \frac{9}{16}}{1 + \cos(\omega) + \frac{1}{4}}$$

Let $N(e^{j\omega})$ be the numerator of the above expression, and $D(e^{j\omega})$ be the denominator of the expression.

$$\begin{aligned} Q(e^{j\omega}) &= \frac{2N(e^{j\omega})}{2N(e^{j\omega}) + M(e^{j\omega})} \\ &= \frac{3\frac{1}{8} + 3\cos(\omega)}{5\frac{3}{8} + 4\cos(\omega)} \end{aligned}$$

(b) In this case, the signal $V[n]$ is not independent of the signal $X[n]$. Therefore, we can not apply the formula derived in part (a). We have to calculate the Wiener filter expression using the general formula, that is:

$$Q(e^{j\omega}) = \frac{S_{XY}(\omega)}{S_Y(\omega)}$$

We need to calculate S_{XY} and S_Y. We will first calculate S_Y in terms of S_{XY} and see how it simplifies the calculations. Let's calculate R_Y in terms of R_{XY}:

$$\begin{aligned} R_Y[m] &= \mathbb{E}\left[Y[n](X[n]+V[n])\right] \\ &= \mathbb{E}\left[Y[n]X[n-m]\right] + \mathbb{E}\left[\frac{1}{3}Y[n]X[n-m-1]\right] - \mathbb{E}\left[\frac{1}{9}Y[n]X[n-m-2]\right] \\ &= R_{XY}[m] + \frac{1}{3}R_{XY}[m+1] - \frac{1}{9}R_{XY}[m+2]. \end{aligned}$$

After taking Fourier transform we get S_Y:

$$S_Y(\omega) = S_{XY}(\omega)\left(1 + \frac{1}{3}e^{j\omega} - \frac{1}{9}e^{2j\omega}\right)$$

And the Wiener filter is given by:

$$Q(e^{j\omega}) = \frac{S_{XY}}{S_Y} = \frac{1}{1 + \frac{1}{3}e^{-j\omega} - \frac{1}{9}e^{-2j\omega}}.$$

Exercise 34.
The process $X[n]$ is a real AR process:

$$X[n] = 0.3X[n-1] - 0.4X[n-2] + 0.5X[n-3] + W[n],$$

where $W[n]$ is a white noise.

(a) What is the order of the above AR process?.

(b) Using Yule Walker equations, give the best linear predictor of order 2 of $X[n]$, i.e. find a and b in $\hat{X}[n] = aX[n-1] + bX[n-2]$ such that the residual $\|\epsilon\|_2^2 = \mathbb{E}[|X[n] - \hat{X}[n]|^2]$ is minimized.

Solution 34.

(a) The AR process is of order 3.

(b) According to the Yule-Walker equations,

$$\begin{bmatrix} R_X[0] & R_X[1] \\ R_X[1] & R_X[0] \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} R_X[1] \\ R_X[2] \end{bmatrix}.$$

Knowing that,

$$\begin{aligned} R_X[1] &= \mathbb{E}[X[n]X[n-1]] \\ &= 0.3\mathbb{E}[X[n-1]X[n-1]] - 0.4\mathbb{E}[X[n-2]X[n-1]] + 0.5\mathbb{E}[X[n-3]X[n-1]] + \mathbb{E}[W[n]X[n-1]] \\ &= 0.3R_X[0] - 0.4R_X[1] + 0.5R_X[2], \end{aligned}$$

and

$$\begin{aligned} R_X[2] &= \mathbb{E}[X[n]X[n-2]] \\ &= 0.3\mathbb{E}[X[n-1]X[n-2]] - 0.4\mathbb{E}[X[n-2]X[n-2]] + 0.5\mathbb{E}[X[n-3]X[n-2]] + \mathbb{E}[W[n]X[n-2]] \\ &= 0.3R_X[1] - 0.4R_X[0] + 0.5R_X[1], \end{aligned}$$

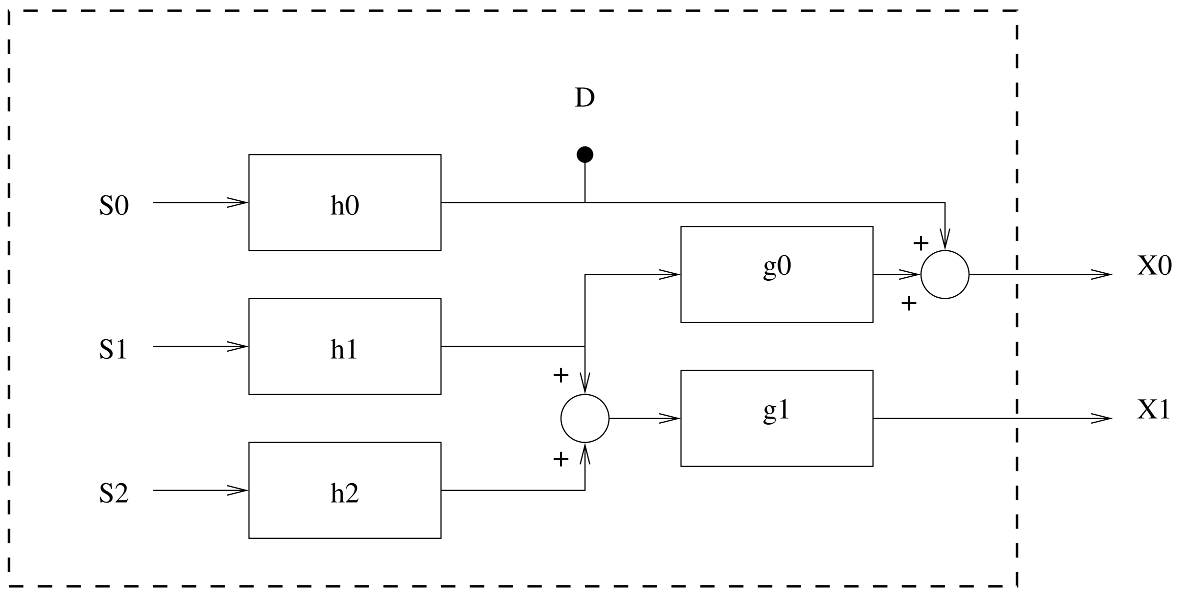
one obtains

$$\begin{aligned} a &\simeq 0.133 \\ b &\simeq -0.333. \end{aligned}$$

Chapter 4

Adaptive Signal Processing

Exercise 35.
Consider the following schematic diagram:



The processes $S_0[n]$, $S_1[n]$, $S_2[n]$ are zero mean white processes, uncorrelated and with variance 1. The filters h_0, h_1, h_2 are causal time invariant linear filters. We know that $H_1(z) = 1 + z^{-1}$, while the filters h_0 and h_2 are unknown. The filter g_0 is also causal and time invariant and the transfer function has the structure

$$G_0(z) = a_0 z^{-1} + a_1 z^{-2} + a_2 z^{-3},$$

where a_0, a_1, a_2 are unknown parameters. The filter g_1 has transfer function $G_1(z) = 1 + z^{-1}$. We want to estimate the process $D[n]$; unfortunately, the elements and the processes inside the rectangle in dashed line are not accessible. Only the processes $X_0[n]$, $X_1[n]$ can be measured and used for the estimation. Some measurements allows to say that the spectral density of the process $X_1[n]$ is given by

$$S_{X_1}(\omega) = 2 + 2\cos\omega.$$

(a) Propose a scheme to estimate $D[n]$ based only on the observation of the processes $X_0[n]$, $X_1[n]$. (State clearly the error process $E[n]$ whose variance is to be minimized by the adaptive filter(s).)

(b) Which is the filter length that you would choose for the adaptive filter(s)? Justify your answer.

(c) Assume that the lenght of the adaptive filter(s) is $L = 4$. What is the range of the step-size that we can consider for the LMS algorithm (steepest-descent range)? Which is a more conservative range?

Solution 35.

(a) The process $D[n]$ can be estimated by the following adaptive filter which uses the least-mean-squares (LMS) algorithm.

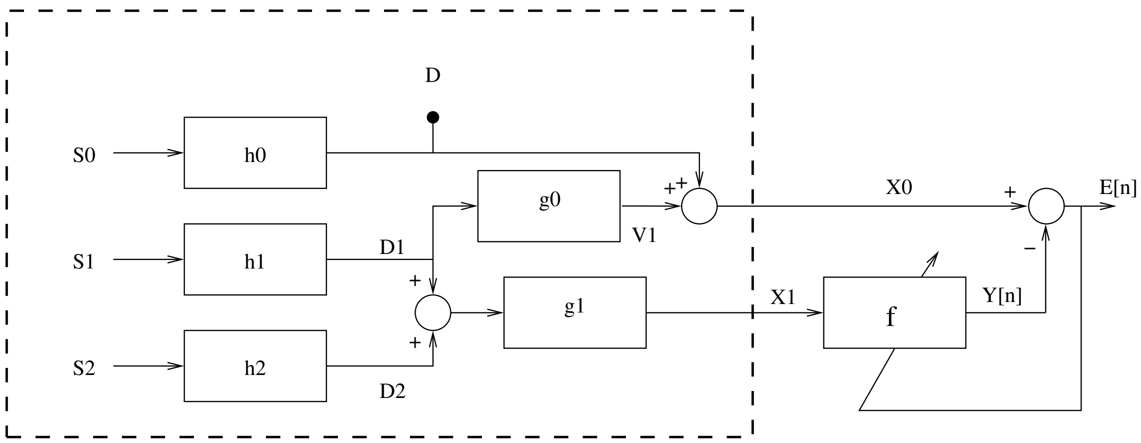


Figure 4.1: Scheme 1.

The cost function to be minimised by the filter is given by

$$\begin{aligned} J_n = \mathbb{E}[E[n]^2] &= \mathbb{E}[(X_0[n] - Y[n])^2] = \mathbb{E}[(D[n] + V[n] - Y[n])^2] \\ &= \mathbb{E}[D[n]^2] + 2\mathbb{E}[D[n](V[n] - Y[n])] + \mathbb{E}[(V[n] - Y[n])^2] \\ &= \mathbb{E}[D[n]^2] + \mathbb{E}[(V[n] - Y[n])^2]. \end{aligned}$$

Note that $D[n]$ and $(V[n] - Y[n])$ are uncorrelated and zero mean processes.

An alternative (and maybe more intuitional) scheme is given in Fig. 4.2, where we first estimate the process $D_1[n]$ by using a Wiener filter and then pass these estimates through an adaptive filter (LMS) which again minimizes the cost function

$$J_n = \mathbb{E}[E[n]^2] = \mathbb{E}[(X_0[n] - Y[n])^2] = \mathbb{E}[D[n]^2] + \mathbb{E}[(V[n] - Y[n])^2]$$

but the input to the adaptive filter is now the estimated process $\hat{D}_1[n]$.

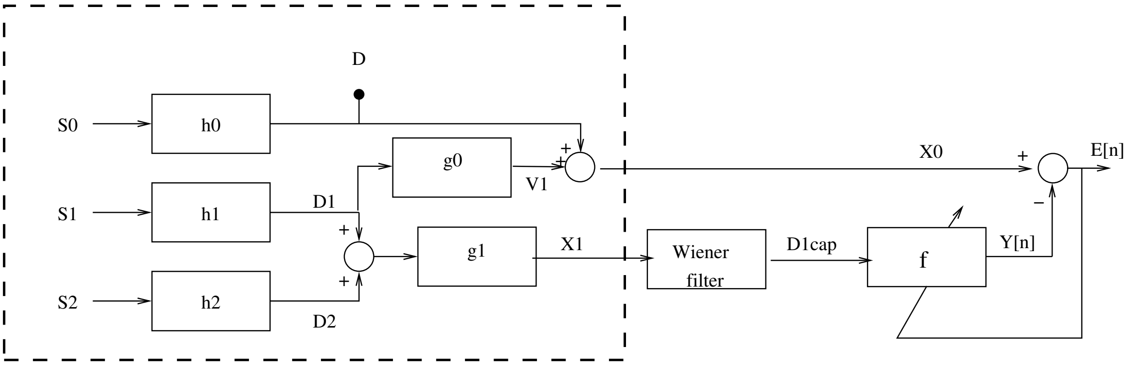


Figure 4.2: Scheme 2.

The Wiener filter is given by

$$W(e^{j\omega}) = \frac{S_{D_1X_1}(\omega)}{S_{X_1}(\omega)}.$$

We know that the spectral density of the process $X_1[n]$ is given by

$$S_{X_1}(\omega) = 2 + 2\cos(\omega) = (1 + e^{j\omega})(1 + e^{-j\omega})$$

and the cross corelation of the processes $D_1[n]$ and $X_1[n]$ can be found as

$$\begin{aligned} R_{D_1X_1}[m] &= \mathbb{E}\left[D_1[n+m]X_1[n]\right] \\ &= \mathbb{E}\left[D_1[n+m]\left(\sum_{k\in\mathbb{Z}} g_1[k](D_1[n-k] + D_2[n-k])\right)\right] \\ &= \sum_{k\in\mathbb{Z}} g_1[k]\left(\mathbb{E}\left[D_1[n+m]D_1[n-k]\right] + \mathbb{E}\left[D_1[n+m]D_2[n-k]\right]\right) \\ &= \sum_{k\in\mathbb{Z}} g_1[k]R_{D_1}[m+k] \\ &= R_{D_1}[m] + R_{D_1}[m+1] \end{aligned}$$

where to obtain the last equality we used the fact that the impulse respose of the filter g_1 is given by $g_1[n] = \delta[n] + \delta[n-1]$. Taking the Fourier transform of this expression and using the filtering formula yields,

$$\begin{aligned} S_{D_1X_1}(\omega) &= (1 + e^{j\omega})S_{D_1}(\omega) \\ &= (1 + e^{j\omega})|H_1(e^{j\omega})|^2 S_{S_1}(\omega) \\ &= (1 + e^{j\omega})^2(1 + e^{-j\omega}). \end{aligned}$$

Hence the Wiener filter is given by

$$W(e^{j\omega}) = \frac{S_{D_1X_1}(\omega)}{S_{X_1}(\omega)} = 1 + e^{j\omega}.$$

In part (b) of the exercise we will see that the optimum solution for the adaptive filter in the first scheme actually decomposes into the structure in the second scheme.

(b) The optimal lenght for the adaptive filters can be determined by looking at the lenght of the optimal solutions for the filters. We concantrate on the first scheme given in Fig. 4.1. Let f^* denote the optimal solution for the adaptive filter in Fig. 4.1 and $F(e^{j\omega})$ denote its transfer function (Fourier transform of f^*). Thus,

$$F(e^{j\omega}) = \frac{S_{X_0X_1}(\omega)}{S_{X_1}(\omega)}$$

and we have

$$\begin{aligned} R_{X_0X_1}[m] &= \mathbb{E}\left[X_0[n+m]X_1[n]\right] \\ &= \mathbb{E}[(D[n+m] + a_0D_1[n+m-1] + a_1D_1[n+m-2] + a_2D_1[n+m-3]) \\ &\cdot (D_1[n] + D_2[n] + D_1[n-1] + D_2[n-1])] \\ &= a_0R_{D_1}[m-1] + a_1R_{D_1}[m-2] + a_2R_{D_1}[m-3] \\ &+ a_0R_{D_1}[m] + a_1R_{D_1}[m-1] + a_2R_{D_1}[m-2]. \end{aligned}$$

Note that $D[n]$, $D_1[n]$ and $D_2[n]$ are uncorrelated processes. Taking the fourier transform,

$$S_{X_0X_1}(\omega) = (1+e^{-j\omega})(a_0 + a_1e^{-j\omega} + a_2e^{-2j\omega})S_{D_1}(\omega).$$

Recalling that $S_{D_1}(\omega) = (1+e^{j\omega})(1+e^{-j\omega}) = S_{X_1}(\omega)$ we obtain

$$\begin{aligned} F(e^{j\omega}) &= (1+e^{-j\omega})(a_0 + a_1e^{-j\omega} + a_2e^{-2j\omega}) && (4.1) \\ &= a_0 + (a_0+a_1)e^{-j\omega} + (a_1+a_2)e^{-2j\omega} + a_2e^{-3j\omega} && (4.2) \\ &= (1+e^{j\omega})(a_0e^{-j\omega} + a_1e^{-2j\omega} + a_2e^{-3j\omega}) && (4.3) \end{aligned}$$

(4.2) shows that the optimal filter f^* has four taps, hence our choice for the length of the adaptive filter in Fig. 4.1 should be $L = 4$, while (4.3) shows that the the optimal solution for the adaptive filter in Fig. 4.1 decomposes in to the structure in Fig. 4.2 since $F(e^{j\omega}) = W(e^{j\omega})G_0(e^{j\omega})$.

(c)

$$R_{X_1}[m] = 2\delta[m] + \delta[m-1] + \delta[m+1]$$

Thus,

$$\boldsymbol{R}_{X_1} = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Using Matlab, one can see that the maximum eigenvalue of $\boldsymbol{R}_{X_1}$ is 3.6180. Hence $0 < \mu < \frac{2}{\lambda_{max}} = 0.5528$. A more conservative range will be $0 < \mu < \frac{2}{L\sigma^2_{X_1}} = 0.25$.

Exercise 36.
Consider the following matlab script

```
hdeterministic = [1 1]';
hrandom = 5;       % number of random terms that have to be identified
hlen = hrandom + length(hdeterministic) - 1; % length of the filter that
we should identify
BigNumber = 2000;   % number of samples that we process
FilterChangeStep = 200; % the filter that we identify changes every
FilterChangeStep samples
eprocess = zeros(BigNumber, 1); % error value for each estimated sample
noisestd = 1e-3; % noise standard deviation
xvector = randn(hlen, 1); % xvector contains the last hlen samples of
the input process

%%% ADD HERE

%initialization of filter f for the first iteration
% (vector of size hlen x 1)
f = zeros(hlen, 1); % this is an example

%%% END ADD

% main loop
for n = 1:BigNumber,
    x = randn(1,1); % new sample of the input process
    xvector = [x; xvector(1:end-1)]; % update xvector
    % check if the filter has to be changed
    if rem(n, FilterChangeStep) == 1,
        h = conv(hdeterministic, rand(hrandom, 1));
    end;
    d = h'*xvector + noisestd * randn(1,1); % desired process
    y = f'*xvector; % current estimate
    e = d - y; % current error

    %%% ADD HERE

    % update of filter f for next iteration
    % YOU ARE NOT ALLOWED TO USE VARIABLE 'h'
    f =

    %%% END ADD

    eprocess(n) = e; % save e for final statistics
end;


figure; plot(eprocess);
fprintf(1, 'The average error variance is %f', mean(eprocess.^2));
```

The goal is to complete the program in order to minimize the variance of the error process (variable "eprocess") by finding an appropriate filter "f" according to the following rules:

- You cannot use the variable "h", which corresponds to the filter that has to be estimated.
- You cannot modify the other variables of the program (only f or your own variables).
- Needless to say, you shall not cheat by playing with the seed of the random number generator!

This is a 'free' exercise, that is, there is no unique correct answer to the problem. One can find different solutions to this exercise, with different complexity and performances. You should try to find the best solution for the problem by using all the information that is available to you on the structure of the identified system.
Do not forget to write comments and if necessary explanations on what your Matlab code does and give a sample output of your program (as well as the code itself).

Solution 36.
Here is one solution we suggest for the problem but you might come up with solutions that perform better than the one below:

```
hdeterministic = [1 1]';
hrandom = 5;       % number of random terms that have to be identified
hlen = hrandom + length(hdeterministic) - 1; % length of the filter that we should identify
BigNumber = 2000;   % number of samples that we process
FilterChangeStep = 200; % the filter that we identify changes every FilterChangeStep samples
eprocess = zeros(BigNumber, 1); % error value for each estimated sample
noisestd = 1e-3; % noise standard deviation
sigmax = 1; % input process standard deviation
xvector = randn(hlen, 1); % xvector contains the last hlen samples of the input process



%%% ADD HERE

%initialization of filter f for the first iteration
% (vector of size hlen x 1)
%f = 0.5 * ones(hlen, 1); % this is an example
frandom =  0.5 * ones(hrandom, 1); f = conv(hdeterministic,
frandom);

%%% END ADD

% main loop
for n = 1:BigNumber,
    x = sigmax * randn(1,1); % new sample of the input process
    xvector = [x; xvector(1:end-1)]; % update xvector
    % check if the filter has to be changed
    if rem(n, FilterChangeStep) == 1,
        h = conv(hdeterministic, rand(hrandom, 1));
    end;
    d = h'*xvector + noisestd * randn(1,1); % observed process
    y = f'*xvector; % current estimate
    e = d - y; % current error

    %%% ADD HERE

    % update of filter f for next iteration
    % YOU ARE NOT ALLOWED TO USE VARIABLE 'h'
```

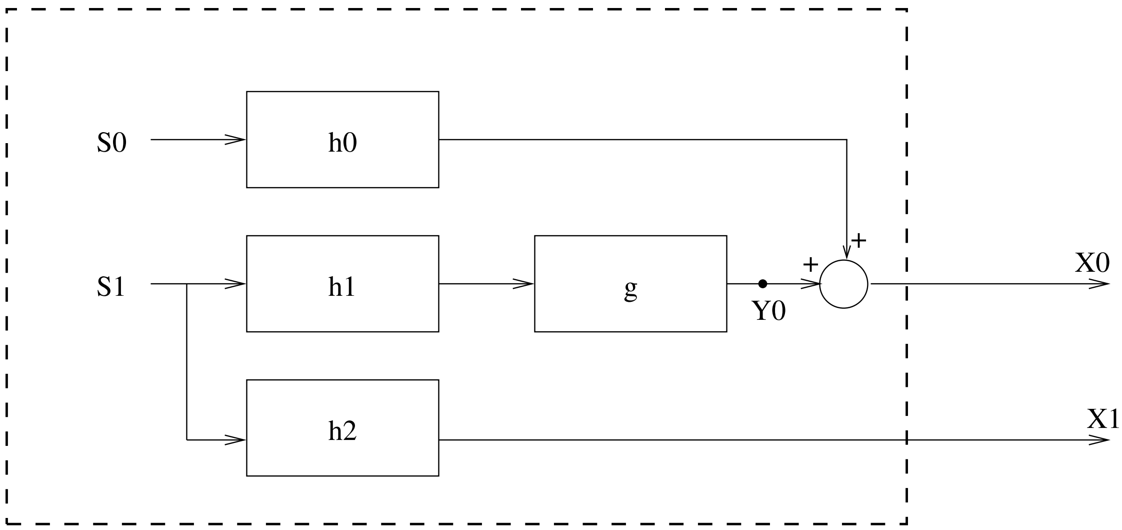
```
    xrandom = conv(hdeterministic, xvector);
    xrandom = xrandom(length(hdeterministic):end-length(hdeterministic)+1);
    mu = 2 / hrandom / (sum(conv(hdeterministic(end:-1:1), hdeterministic)) * sigmax^2) / 2;
    % mu = 2 / L / S_max / 2
    frandom = frandom + mu * e * xrandom;
    if rem(n + 1, FilterChangeStep) == 1,
        frandom =  0.5 * ones(hrandom, 1);
    end;
    f = conv(hdeterministic, frandom);
    % f = f + 0.1 * e * xvector; % this is an example

    %%% END ADD

    eprocess(n) = e; % save e for final statistics
end;

figure; plot(eprocess);
fprintf(1, 'The average error variance is %f', mean(eprocess.^2));
```

Exercise 37.
Consider the following diagram:



The processes $S_0[n]$, $S_1[n]$ are jointly Gaussian, uncorrelated, white with zero-mean and unit variance. The filters h_0, h_1, h_2 have z-transforms

$$H_0(z) = 1 - z^{-1}, \qquad H_1(z) = 1 + z^{-1}, \qquad H_2(z) = 1 + \frac{1}{2}z^{-1}$$

respectively. The filter g is not completely specified. We only know that:

1) It is stable.
2) The z-transform of g has the form

$$G(z) = \frac{1}{a_0 + a_1 z^{-}1 + a_2 z^{-2}}$$

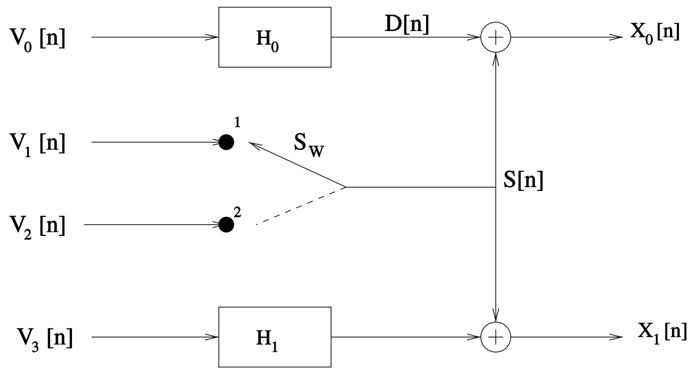
where a_0, a_1, a_2 are unknown real constants.

Answer precisely the following questions:

1) Are the processes $X_0[n]$, $X_1[n]$ Gaussian? Are they stationary?
2) Propose an algorithm that, by using only the measurements of the processes X_0 and X_1, estimates the coefficients a_0, a_1, a_2 of $G(z)$. Remark: You should give enough information to be able to write a computer program, for example use a "pseudo code". For the adaptive elements, if any, computation of the step size will be addressed in the next question.
3) Assume by some means that we are able to estimate that the variance of the process Y, σ_Y^2 is in the range $1 < \sigma_Y^2 < 2$. In such a case, what is a reasonable range for the step-size of the adaptive elements used in the previous answer?
4) Consider the maximum step-size μ^* computed in the previous question. Compare the behavior of the algorithm when we take a step size $\mu^*/2$ and $\mu^*/10$. (Explain in words how the behavior of the algorithm differs when we use these two different values for the step-size.)

Solution 37.

Exercise 38.
Consider the following diagram:



where H_0 and H_1 are causal filters with the transfer functions:

$$H_0(z) = 1 + z^{-1},$$

$$H_1(z) = 1 - z^{-1}.$$

V_0, V_1, V_2 and V_3 are white stationary processes, uncorrelated, jointly gaussian with zero mean and the variances

$$\sigma^2_{V_0} = 1 \quad \sigma^1_{V_1} = 1 \quad \sigma^2_{V_2} = 2 \quad \sigma^2_{V_3} = 1.$$

The switch S_W is in the position 1 when the time index n is even and 2 when n is odd.

(a) Is $X_0[n]$ Gaussian process? Is $X_0[n]$ wide sense stationary process? Compute the correlation of $X_0[n]$.

(b) Determine the optimal filter that estimates the process $D[n]$, i.e. $\hat{D}[n]$, given the observation of the two samples $X_0[n]$ and $X_0[n-1]$.

(c) Determine the optimal filter that estimates the process $S[n]$, i.e. $\hat{S}[n]$, given the observation of the two samples $X_1[n]$ and $X_1[n-1]$.

How can we use such an estimator to determine an estimate of $D[n]$? Compare this solution with the one obtained in question 2), which one would you prefer?

(d) Consider the output of the system as a vector $\bar{X}[n] = [X_0[n]X_1[n]]^T$ and give the expression for the optimal estimator of $D[n]$ based on the two observation of $\bar{X}[n]$ and $\bar{X}[n-1]$.

Hint: Write the estimator as

$$\hat{D}[n] = \bar{f}^T_{n,0}\bar{X}[n] + \bar{f}^T_{n,1}\bar{X}[n-1].$$

Recall:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - cb}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Solution 38.

(a) We have:

$$X_0[n] = V_0[n] + V_0[n-1] + S[n] = \\ = \begin{cases} V_0[n] + V_0[n-1] + V_1[n], & \text{n is even} \\ V_0[n] + V_0[n-1] + V_2[n], & \text{n is odd} \end{cases}$$

Process $X_0[n]$ is a sum of two processes $D[n]$ and $S[n]$ that are Gaussian at every instant. Therefore, $X_0[n]$ is Gaussian, as well. To check if the process is wide sense stationary we compute the mean and the variance.

$$\mathbb{E}[X_0[n]] = 0$$

$$\mathbb{E}[X_0[n]X_0[n]] = \mathbb{E}[(V_0[n] + V_0[n-1] + S[n])(V_0[n] + V_0[n-1] + S[n])] \\ = \begin{cases} 2\sigma_{V_0}^2 + \sigma_{V_1}^2, & \text{n is even} \\ 2\sigma_{V_0}^2 + \sigma_{V_2}^2, & \text{n is odd} \end{cases}$$

We can see that $S[n]$ is not a wss process and consequently $X_0[n]$ is not wss.

In order to compute the correlation of $X_0[n]$ we need first to compute the correlation of $S[n]$.

$$R_S[n,m] = \mathbb{E}[S[n]S[m]] = \begin{cases} \mathbb{E}[V_1[n]V_1[m]] = \delta[n-m]\sigma_{V_1}^2, & \text{n and m are even} \\ \mathbb{E}[V_1[n]V_2[m]] = 0, & \text{n even, m odd} \\ \mathbb{E}[V_2[n]V_1[m]] = 0, & \text{n odd, m even} \\ \mathbb{E}[V_2[n]V_2[m]] = \delta[n-m]\sigma_{V_2}^2, & \text{n and m are odd} \end{cases}$$

Then,

$$\begin{aligned} R_{X_0}[n,m] &= \mathbb{E}[X_0[n]X_0[m]] \\ &= \mathbb{E}[(V_0[n] + V_0[n-1] + S[n])(V_0[m] + V_0[m-1] + S[m])] \\ &= 2\sigma_{V_0}^2\delta[n-m] + \sigma_{V_0}^2\delta[n-1-m] + \sigma_{V_0}^2\delta[n-m+1] + R_S[n,m] \\ &= \begin{cases} 2\sigma_{V_0}^2\delta[n-m] + \sigma_{V_1}^2\delta[n-m], & \text{n and m are even} \\ 2\sigma_{V_0}^2\delta[n-m] + \sigma_{V_2}^2\delta[n-m], & \text{n and m are odd} \\ \sigma_{V_0}^2\delta[n-1-m] + \sigma_{V_0}^2\delta[n-m+1], & \text{otherwise} \end{cases} \end{aligned}$$

(b) We define the cost function as:

$$J_{min} = \mathbb{E}[|D[n] - \sum_{k=0}^{1} f_n(k)X_0[n-k]|^2].$$

The optimal filter is given by:

$$\begin{bmatrix} f_n(0) \\ f_n(1) \end{bmatrix} = \begin{bmatrix} R_{X_0}[n,n] & R_{X_0}[n-1,n] \\ R_{X_0}[n,n-1] & R_{X_0}[n-1,n-1] \end{bmatrix}^{-1} \begin{bmatrix} R_{DX_0}[n,n] \\ R_{DX_0}[n,n-1] \end{bmatrix}$$

where

$$R_{DX_0}[n,m] = \mathbb{E}[D[n]X_0[m]] = \mathbb{E}[D[n](D[m] + S[m])] = R_D[n,m] + R_{DS}[n,m],$$

$$R_D[n,m] = 2\sigma_{V_0}^2\delta[n-m] + \sigma_{V_0}^2\delta[n-1-m] + \sigma_{V_0}^2\delta[n-m+1],$$

$$R_{DS}[n,m] = 0 \quad \text{for all n and m}$$

Then, when n is even we have:

$$\begin{bmatrix} f_n(0) \\ f_n(1) \end{bmatrix} = \begin{bmatrix} 2\sigma^2_{V_0} + \sigma^2_{V_1} & \sigma^2_{V_0} \\ \sigma^2_{V_0} & 2\sigma^2_{V_0} + \sigma^2_{V_2} \end{bmatrix}^{-1} \begin{bmatrix} 2\sigma^2_{V_0} \\ \sigma^2_{V_0} \end{bmatrix} = \begin{bmatrix} 7/11 \\ 1/11 \end{bmatrix},$$

and when n is odd we have:

$$\begin{bmatrix} f_n(0) \\ f_n(1) \end{bmatrix} = \begin{bmatrix} 2\sigma^2_{V_0} + \sigma^2_{V_2} & \sigma^2_{V_0} \\ \sigma^2_{V_0} & 2\sigma^2_{V_0} + \sigma^2_{V_1} \end{bmatrix}^{-1} \begin{bmatrix} 2\sigma^2_{V_0} \\ \sigma^2_{V_0} \end{bmatrix} = \begin{bmatrix} 5/11 \\ 2/11 \end{bmatrix}.$$

(c) We define the cost function as:

$$J_{min} = \mathbb{E}[|S[n] - \sum_{k=0}^{1} f_n(k) X_1[n-k]|^2].$$

The optimal filter is given by:

$$\begin{bmatrix} f_n(0) \\ f_n(1) \end{bmatrix} = \begin{bmatrix} R_{X_1}[n,n] & R_{X_1}[n-1,n] \\ R_{X_1}[n,n-1] & R_{X_1}[n-1,n-1] \end{bmatrix}^{-1} \begin{bmatrix} R_{SX_1}[n,n] \\ R_{SX_0}[n,n-1] \end{bmatrix}$$

where

$$\begin{aligned} R_{X_1}[n,m] &= \mathbb{E}[X_1[n]X_1[m]] \\ &= \mathbb{E}[(V_3[n] - V_3[n-1] + S[n])(V_3[m] - V_3[m-1] + S[m])] \\ &= 2\sigma^2_{V_3}\delta[n-m] - \sigma^2_{V_3}\delta[n-1-m] - \sigma^2_{V_3}\delta[n-m+1] + R_S[n,m] \\ &= \begin{cases} 2\sigma^2_{V_3}\delta[n-m] + \sigma^2_{V_1}\delta[n-m], & \text{n and m are even} \\ 2\sigma^2_{V_3}\delta[n-m] + \sigma^2_{V_2}\delta[n-m], & \text{n and m are odd} \\ -\sigma^2_{V_3}\delta[n-1-m] - \sigma^2_{V_3}\delta[n-m+1], & \text{otherwise} \end{cases}, \end{aligned}$$

$$\begin{aligned} R_{SX_1}[n,m] = \mathbb{E}[S[n]X_1[m]] &= \mathbb{E}[S[n](V_3[m] - V_3[m-1] + S[m])] \\ = R_S[n,m] &= \begin{cases} \sigma^2_{V_1}\delta[n-m] & \text{n,m even} \\ \sigma^2_{V_2}\delta[n-m] & \text{n,m odd} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

For even n we have:

$$\begin{bmatrix} f_n(0) \\ f_n(1) \end{bmatrix} = \begin{bmatrix} 2\sigma^2_{V_3} + \sigma^2_{V_1} & -\sigma^2_{V_3} \\ -\sigma^2_{V_3} & 2\sigma^2_{V_3} + \sigma^2_{V_2} \end{bmatrix}^{-1} \begin{bmatrix} \sigma^2_{V_1} \\ 0 \end{bmatrix} = \begin{bmatrix} 6/17 \\ 1/17 \end{bmatrix},$$

and when n is odd we have:

$$\begin{bmatrix} f_n(0) \\ f_n(1) \end{bmatrix} = \begin{bmatrix} 2\sigma^2_{V_3} + \sigma^2_{V_2} & -\sigma^2_{V_3} \\ -\sigma^2_{V_3} & 2\sigma^2_{V_3} + \sigma^2_{V_1} \end{bmatrix}^{-1} \begin{bmatrix} \sigma^2_{V_2} \\ 0 \end{bmatrix} = \begin{bmatrix} 12/17 \\ 2/17 \end{bmatrix}.$$

The estimated of $D[n]$ can be determine as:

$$\hat{D}[n] = X_0[n] - \hat{S}[n]$$

(d) The optimal filter is obtained in the same way as before

$$\begin{bmatrix} \bar{f}_{n,0} \\ \bar{f}_{n,1} \end{bmatrix} = \begin{bmatrix} R_{\bar{X}}[n,n] & R_{\bar{X}}[n-1,n] \\ R_{\bar{X}}[n,n-1] & R_{\bar{X}}[n-1,n-1] \end{bmatrix}^{-1} \begin{bmatrix} R_{D\bar{X}}[n,n] \\ R_{D\bar{X}}[n,n-1] \end{bmatrix}$$

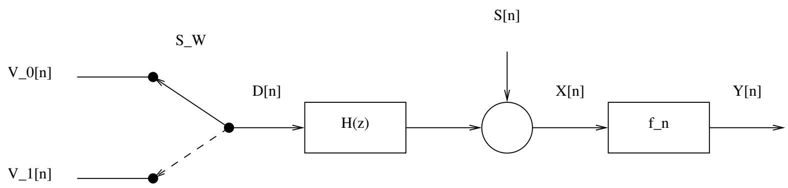
where

$$R_{\bar{X}}[n,m] = R\left[\begin{bmatrix} X_0[n] \\ X_1[n] \end{bmatrix} \begin{bmatrix} X_0^*[m] & X_1^*[m] \end{bmatrix}\right] = \begin{bmatrix} R_{X_0}[n,m] & R_{X_0X_1}[n,m] \\ R_{X_1X_0}[n,m] & R_{X_1}[n,m] \end{bmatrix},$$

and

$$R_{D\bar{X}}[n,m] = \begin{bmatrix} R_{DX_0}[n,m] \\ R_{DX_1}[n,m] \end{bmatrix}.$$

Exercise 39.
Consider the following schematic diagram:



The two processes $V_0[n]$ and $V_1[n]$ are zero mean jointly Gaussian, they are mutually uncorrelated and their self correlation functions are

$$R_{V_0}[n] = \mathbb{E}[V_0[n+m]V_0[m]] = \rho_0^{|n|},$$

$$R_{V_1}[n] = \mathbb{E}[V_1[n+m]V_1[m]] = \rho_1^{|n|}.$$

In the following we will take $\rho_0 = 1/2$, $\rho_1 = 1/3$. The switch S_W selects one of the two processes to generate the "desired" process $D[n]$ that has to be estimated. The measurements are obtained by filtering $D[n]$ with the filter $H(z) = 1 + z^{-1}$ and adding the noise process $S[n]$, which is i.i.d, zero mean, jointly Gaussian with V_0 and V_1 and with variance $\sigma_S^2 = 1$. The measurement process $X[n]$ is filtered by the time-varying filter f_n of length $L = 3$ to obtain the estimate $Y[n]$. The goal is to minimize the variance of the estimation error $E[n] = D[n] - Y[n]$.

(a) Assume first that the switch is in the position "0" (i.e. $D[n] = V_0[n]$). Write the normal equations for the filter f_n and find the optimal linear filter. Is this a Wiener filter? Compute the estimation error variance, $\mathbb{E}[E[n]^2]$. Do the same with the switch in the position "1".

(b) Assume now that the switch is in the position "0" for the even samples and "1" for the odd samples. Do the following steps:

   (a) Compute the correlation function $R_D[n,m] = \mathbb{E}[D[n]D[m]]$. Is the process $D[n]$ stationary?

   (b) Compute the correlation functions $R_X[n,m] = \mathbb{E}[X[n]X[m]]$ and $R_{DX}[n,m] = \mathbb{E}[D[n]X[m]]$.

   (c) Write the normal equations for the even and odd time indexes. Find the optimal linear filter and the error variance for the two cases (even and odd time indexes) and compare them with the result of question a).

(c) Assume that the position of the switch is chosen randomly and independently for each sample. The probability of position "0" is $p_0 = 1/2$. Compute again the correlation $R_D[n,m] = \mathbb{E}[D[n]D[m]$ and check if the process is stationary. Compute the optimal linear filter in this case and compare the answer with the results of question b).

Solution 39.

(a) We want to minimize the error

$$E[n] \quad = \quad D[n] - Y[n] = D[n] - \sum_i f_n[i]X[n-i]$$

where

$$X[n] = \sum_k h[k]D[n-k] + S[n].$$

If we define the cost function as

$$J_n = \mathbb{E}[E[n]^2],$$

than it has a unique minimum which can be found by setting the first derivative to zero. Following the steps that are given in the section 4.2.1 of the lecture notes, we find that the *normal* equation is:

$$\sum_{j=0}^{L-1} f_n[j]R_X[n-j,n-i] = R_{DX}[n,n-i] \quad i=0,\ldots,L-1, \quad \forall n \in \mathbb{Z}.$$

Now, we compute $R_X[n-j,n-i]$ and $R_{DX}[n,n-i]$ for the case where $D[n] = V_0[n]$.

$$\begin{aligned} R_X[n-j,n-i] &= \mathbb{E}[(h[0]V_0[n-j] + h[1]V_0[n-j-1] + S[n-j]) \\ &\quad (h[0]V_0[n-i] + h[1]V_0[n-i-1] + S[n-i])] \\ &= 2R_{V_0}[i-j] + R_{V_0}[i-j+1] + R_{V_0}[i-j-1] + \sigma_S^2\delta[i-j] \\ &= 2\rho_0^{|i-j|} + \rho_0^{|i-j+1|} + \rho_0^{|i-j-1|} + \sigma_S^2\delta[i-j], \end{aligned}$$

$$\begin{aligned} R_{DX}[n,n-i] &= \mathbb{E}[V_0[n] \cdot ((h[0]V_0[n-i] + h[1]V_0[n-i-1] + S[n-i]) \\ &= R_{V_0}[i] + R_{V_0}[i+1] \\ &= \rho_0^{|i|} + \rho_0^{|i+1|}, \end{aligned}$$

We can see that the processes $D[n]$ and $X[n]$ are stationary BUT the filter f_n is not a Wiener filter since we limit the length of the filter to be $L = 3$.

From the Yule-Walker equation we have

$$f_n = R_{X,n}^{-1} r_{DX,n}.$$

$$\begin{bmatrix} f_n[0] \\ f_n[1] \\ f_n[2] \end{bmatrix} = \begin{bmatrix} 2+2\rho_0+1 & 2\rho_0+1+\rho_0^2 & 2\rho_0^2+\rho_0+\rho_0^3 \\ 2\rho_0+1+\rho_0^2 & 2+2\rho_0+1 & 2\rho_0+\rho_0^2+1 \\ 2\rho_0^2+\rho_0+\rho_0^3 & 2\rho_0+1+\rho_0^2 & 2+2\rho_0+1 \end{bmatrix}^{-1} \begin{bmatrix} 1+\rho_0 \\ \rho_0+\rho_0^2 \\ \rho_0^2+\rho_0^3 \end{bmatrix}.$$

Changing the $\rho_0 = 1/2$ we get

$$f_n = [0.3944 \quad -0.0361 \quad 0.0031]^T.$$

We compute $\mathbb{E}[|E[n]|^2]$ from the formula:

$$\begin{aligned} \mathbb{E}[|E[n]|^2] &= \mathbb{E}[(D[n] - f^T X_n)^2] \\ &= \sigma_D^2 + f^T R_X f - 2f^T r_{DX} \end{aligned}$$

Then, we can find

$$\mathbb{E}[|E[n]|^2] = 0.4343.$$

When the switch is in the position "1", all the steps are the same and we need to change V_0 to V_1. In that case we have,

$$\begin{aligned}\begin{bmatrix} f_n[0] \\ f_n[1] \\ f_n[2] \end{bmatrix} &= \begin{bmatrix} 2+2\rho_1+1 & 2\rho_1+1+\rho_1^2 & 2\rho_1^2+\rho_1+\rho_1^3 \\ 2\rho_1+1+\rho_1^2 & 2+2\rho_1+1 & 2\rho_1+\rho_1^2+1 \\ 2\rho_1^2+\rho_1+\rho_1^3 & 2\rho_1+1+\rho_1^2 & 2+2\rho_1+1 \end{bmatrix}^{-1} \begin{bmatrix} 1+\rho_1 \\ \rho_1+\rho_1^2 \\ \rho_1^2+\rho_1^3 \end{bmatrix} \\ &= \begin{bmatrix} 0.3999 \\ -0.0797 \\ 0.0144 \end{bmatrix}\end{aligned}$$

and

$$\mathbb{E}[|E[n]|^2] = \mathbb{E}[(D[n] - f^T X_n)^2] = 0.4912.$$

(b) - We need to distinguish the four cases:

$$R_D[n,m] = \begin{cases} \rho_0^{|n-m|} & \text{n, m even,} \\ \rho_1^{|n-m|} & \text{n, m odd,} \\ 0 & \text{n even, m odd,} \\ 0 & \text{n odd, m even.} \end{cases}$$

Clearly, the process $D[n]$ is not stationary.

- Let us call $D_h[n] = \sum_k h[k]D[n-k] = D[n] - D[n-1]$. Then

$$\begin{aligned} R_X[n,m] &= \mathbb{E}[X[n]X[m]] = \mathbb{E}[(D_h[n] + S[n])(D_h[m] + S[m])] \\ &= R_{D_h}[n,m] + R_S[n,m] = R_{D_h}[n,m] + \sigma_S^2 \delta[n-m] \end{aligned}$$

and

$$\begin{aligned} R_{D_h}[n,m] &= \mathbb{E}[(D[n] + D[n-1])(D[m] + D[m-1])] \\ &= R_D[n,m] + R_D[n,m-1] + R_D[n,m-1] \\ &\quad + R_D[n-1,m] + R_D[n-1,m-1] \\ &= \begin{cases} \rho_0^{|n-m|} + 0 + 0 + \rho_1^{|n-m|} & \text{n, m even,} \\ \rho_1^{|n-m|} + 0 + 0 + \rho_0^{|n-m|} & \text{n, m odd,} \\ 0 + \rho_0^{|n-m+1|} + \rho_1^{|n-m-1|} + 0 & \text{n even, m odd,} \\ 0 + \rho_1^{|n-m+1|} + \rho_0^{|n-m-1|} + 0 & \text{n odd, m even.} \end{cases} \end{aligned}$$

The correlation R_{DX} is equal to

$$\begin{aligned} R_{DX}[n,m] &= \mathbb{E}[D[n]X[m]] = \mathbb{E}[D[n](D[m] + D[m-1] + S[m])] \\ &= R_D[n,m] + R_D[n,m-1] \\ &= \begin{cases} \rho_0^{|n-m|} + 0 & \text{n, m even,} \\ \rho_1^{|n-m|} + 0 & \text{n, m odd,} \\ 0 + \rho_0^{|n-m+1|} & \text{n even, m odd,} \\ 0 + \rho_1^{|n-m+1|} & \text{n odd, m even.} \end{cases} \end{aligned}$$

Clearly, the process is not stationary.

- We have the normal equation

$$\sum_{j=0}^{L-1} f_n[j] R_X[n-j, n-i] = R_{DX}[n, n-i] \quad i = 0, \ldots, L-1, \quad \forall n \in \mathbb{Z}$$

and to evaluate R_X and R_{DX}, we need to consider the cases when n is even and n is odd.
Let us, for example, consider the case when n is even. Then,

$$\begin{aligned}
\begin{bmatrix} f_n[0] \\ f_n[1] \\ f_n[2] \end{bmatrix} &= \begin{bmatrix} R_X[n,n] & R_X[n-1,n] & R_X[n-2,n] \\ R_X[n,n-1] & R_X[n-1,n-1] & R_X[n-2,n-1] \\ R_X[n,n-2] & R_X[n-1,n-2] & R_X[n-2,n-2] \end{bmatrix}^{-1} \begin{bmatrix} R_{DX}[n,n] \\ R_{DX}[n,n-1] \\ R_{DX}[n,n-2] \end{bmatrix} \\
&= \begin{bmatrix} 3 & 1+\rho_0^2 & \rho_0^2+\rho_1^2 \\ \rho_0^2+1 & 3 & 1+\rho_1^2 \\ \rho_0^2+\rho_1^2 & \rho_1^2+1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \rho_0^2 \\ \rho_0^2 \end{bmatrix}
\end{aligned}$$

and we compute

$$f_n = [0.3644 \quad -0.0963 \quad 0.0752]^T.$$

Applying the same formula for computing the error as in the previous part, we get:

$$\mathbb{E}[|E[n]|^2] = 0.6409$$

The process $D[n]$ is not stationary and this explains why the error is larger than in the part a) for both $D[n] = V_0[n]$ and $D[n] = V_1[n]$.

(c) Since the position of the switch is randomly chosen we can introduce the random variable $S_W[n]$ that describe the position of the switch. Positions $p0$ and $p1$ appear with the same probability of 1/2. To compute $R_D[n,m] = \mathbb{E}[D[n]D[m]]$, we can use the following formula:

$$\begin{aligned}
\mathbb{E}[f(D)] &= \mathbb{E}[\mathbb{E}[f(D)|S_W]] \\
&= \frac{1}{4}\mathbb{E}[f(D)|s_W = (0,0)] + \frac{1}{4}\mathbb{E}[f(D)|s_W = (0,1)] \\
&\quad + \frac{1}{4}\mathbb{E}[f(D)|s_W = (1,0)] + \frac{1}{4}\mathbb{E}[f(D)|s_W = (1,1)]\,.
\end{aligned}$$

Then

$$\begin{aligned}
R_D[n,m] &= \mathbb{E}[D[n]D[m]] \\
&= \frac{1}{4}\mathbb{E}[D[n]D[m]|s_W = (0,0)] + \frac{1}{4}\mathbb{E}[D[n]D[m]|s_W = (1,1)] \\
&= \frac{1}{4}\rho_0^{|n-m|} + \frac{1}{4}\rho_1^{|n-m|}
\end{aligned}$$

The process is stationary.

To compute the optimal filter we need:

$$\begin{aligned}
R_X[n,m] &= \mathbb{E}[X[n]X[m]] = \mathbb{E}[(D[n] + D[n-1] + S[n])(D[m] + D[m-1] + S[m])] \\
&= 2R_D[n-m] + R_D[n-m+1] + R_D[n-m-1] + \sigma_S^2 \delta[n-m],
\end{aligned}$$

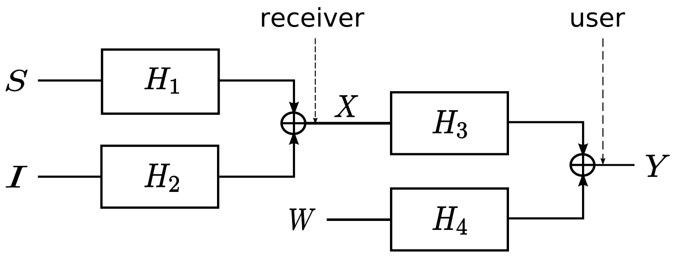
and

$$\begin{aligned} R_{DX}[n,m] &= \mathbb{E}[D[n]X[m]] = \mathbb{E}[D[n](D[m] + D[m-1] + S[m])] \\ &= R_D[n-m] + R_D[n, m-1]. \end{aligned}$$

Changing in the normal equation we get:

$$\begin{bmatrix} f[0] \\ f[1] \\ f[2] \end{bmatrix} = \begin{bmatrix} 2.4167 & 1.0069 & 0.4294 \\ 1.0069 & 2.4167 & 1.0069 \\ 0.4294 & 1.0069 & 2.5167 \end{bmatrix}^{-1} \begin{bmatrix} 0.7083 \\ 0.2986 \\ 0.1308 \end{bmatrix} = \begin{bmatrix} 0.2923 \\ 0.0010 \\ 0.0018 \end{bmatrix}$$

Exercise 40. Wiener and Adaptive Filters
Consider the system in the figure below.



Signals $S[n]$, $I[n]$ and $W[n]$ are centered (zero-mean), i.i.d., jointly Gaussian and uncorrelated random processes, with variances σ_S^2, σ_I^2 and σ_W^2, respectively. Furthermore, filters $H_1(z)$, $H_2(z)$, $H_3(z)$ and $H_4(z)$ have the following forms:

$$\begin{aligned} H_1(z) &= 1 + a_1 z^{-1} \\ H_2(z) &= 1 + a_2 z^{-1} \\ H_3(z) &= 1 + a_3 z^{-1} \\ H_4(z) &= 1 + a_4 z^{-1} . \end{aligned}$$

1 Give the power spectral densities of processes $X[n]$ and $Y[n]$.

2 Assuming you are given access to the interfering signal $I[n]$, show how you would connect an adaptive filter to remove the interfering signal $I[n]$ from the user signal $Y[n]$. What would be the length of the adaptive filter?

3 *(Comment)* What would be needed for estimating the signal $S[n]$ from the signal $Y[n]$ optimally in the mean squared error sense?

Solution 40.

Chapter 5

Spectral Estimation

Exercise 41. AR once again
Consider a centered (zero-mean) real-valued AR process $\{X_n\}_{n\in\mathbb{Z}}$ verifying the equation

$$X[n+1] = aX[n] + W[n+1], \quad n \in \mathbb{Z}$$

where

- $a \in \mathbb{R}$, $|a| < 1$,
- $W[n]$ is a real-valued white noise (*i.e.*, a sequence of i.i.d. random variables), centered, with variance $\sigma^2 > 0$.

We now observe a realization $x[n]$ of the AR process $X[n]$ and we would like to estimate the power spectral density.

(a) Describe precisely a parametric method for estimating the power spectral density of the AR process $X[n]$.

(b) Suppose you can choose to observe either 100 or 1000 realizations of $X[n]$. How many realizations would you choose for your spectral estimator? Justify your answer precisely.

(c) Propose a recursive method for estimating the power spectral density of a more general AR process

$$X[n+1] = \sum_{i=0}^{N-1} a_i X[n-i] + W[n+1], \quad n \in \mathbb{Z}\,.$$

(d) Compare the computational burden of the recursive method with the one of the direct approach.

Solution 41. AR once again

(a) Using the symbolic notation we can express the process $X[n]$ as

$$X[n+1] = aX[n] + W[n+1],$$

$$X[n+1](1 - az^{-1}) = W[n+1],$$

$$X[n+1] = \frac{1}{1 - az^{-1}} W[n+1].$$

Then, the power spectral density $S_X(\omega)$ is given as:

$$S_X(\omega) = \frac{1}{|1 - ae^{-j\omega}|^2}\sigma^2 = \frac{1}{1 + a^2 - 2a\cos\omega}\sigma^2.$$

We need to estimate a and ω^2. The two parameters can be estimated using Yule-Walker equations.

(b) Since the noise is white and centered then it is always better to use more realizations for estimating the covariance matrix used in Yule-Walker equations.

(c) The power spectral density of a more general AR process can be estimated by first estimating the parameters $a_0, \ldots, a_{N-1}$ with Levinson's algorithm, by starting with a one-step predictor, and iteratively computing the coefficients of higher-order predictors until the coefficients of the N-th order predictor has been determined.

(d) In the case of having a model whose order N is known a priory, the computational complexities of both using Levinson's recursive algorithm and directly solving Yule-Walker equations are $\mathcal{O}(N^2)$. However, in the case where the model order N is not known a priory, the computational complexity of Levinson's recursive algorithm stays the same, whereas the complexity of iteratively solving Yule-Walker equations for different orders n, until the right order N has been found, is $\mathcal{O}(N^3)$.

Exercise 42. ANNIHILATING FILTER METHOD VS. MUSIC
Assume that we have a random process $X[n]$ that is composed of 3 complex sinusoids:

$$X[n] = \sum_{k=1}^{3} \alpha_k e^{j(2\pi f_k n + \Theta_k)},$$

where $(f_1, f_2, f_3) = (0.2, 0.3, 0.4)$, $(\alpha_1, \alpha_2, \alpha_3) = (1, 2, 3)$ and the phases Θ_k are stationary random variable, independent and uniformly distributed over $[0, 2\pi)$. The signal is affected by additive zero-mean white noise with σ_W^2, independent of $X[n]$. We have the access only to the noisy realizations, i.e.

$$Y[n] = X[n] + W[n].$$

(a) Simulate 20 realizations of $Y[n]$ when $\sigma_W^2 = 1$ and from this realizations estimate the frequencies f_k and the weights α_k of the sinusoids using:

  (a) annihilating filter method,

  (b) MUSIC method.

(b) Do the same steps when $\sigma_W^2 = 4$ and compare the two methods.

(c) Assume that the signal $X[n]$ is deterministic, i.e. the phases Θ_k are known. We want to estimate f_k and α_k. Can we now use the annihilating filter method and the MUSIC method? Point out the differences.

Solution 42. ANNIHILATING FILTER VS. MUSIC

(a) In Matlab we have the following code:

```
% Signal X
c = [ 1 2 3]; f = [.2 .3 .4]; X = zeros(1,30);
% We have 20 realizations of the process
for i = 1:20
    theta = 2*pi*rand(1,3);
    W = 1*randn(1); % or 4*randn(1);
    % We choose 30 samples from each realization
    n = 0:29; X(i,:) = c.*exp(j*theta)*exp(j*2*pi*f'*n) + W;
end

% ANNIHILATING FILTER METHOD

% In this part we can only use one realization
Xl = toeplitz(conj(X(1,3:29)'), X(1,3:-1:1));

Xr = -conj(X(1,4:30)');

h = pinv(Xl)*Xr; root = roots([1conj(h')]);

% frequency estimates
fe = phase(root)/2/pi
```

```
% weight estimates
n1 = [ 0 1 2]'; ce = abs(inv(exp(j*2*pi*n1*fe')) *conj(X(1,1:3)'))

% As the result we have that the frequency the sinusoid with the amplitude
% 1 and frequency 0.2 is not well estimated since it is usually barried
% into the noise. The estimation accuracy for the weights are usually very
% poor.

% MUSIC

% First, we estimate the covariance matrix.
R = zeros(5,5);
for p = 1:20    % 20 realizations
    for k=5:30  % 30 samples
        R = R + conj([X(p,k) X(p,k-1) X(p,k-2) X(p,k-3) X(p,k-4)])'*
         [X(p,k)' X(p,k-1)' X(p,k-2)' X(p,k-3)' X(p,k-4)'];
    end
end
R = R/26/20; % average over the number of samples {1/(M-N)} and number of realizations
[G S V] = svd(R);
Gnoise=G(:,4:5); % the eigenvectors that correspond to the noise space

% plot the function in 100 points
n2 = [ 0 1 2 3 4 ]; for k = 1:100
    root_music(k)=exp(j*2*pi*(k-1)*n2/100)*Gnoise*Gnoise'*exp(j*2*pi*(k-1)*n2/100)';
    music(k) = 1/real(root_music(k));
end

figure; plot([0:1/100:1-1/100], real(root_music));
% The frequencies of the zeros of the plot correspond to the frequencies of
% the signal X.
figure; plot([0:99], music);
% The frequencies of the peaks of the plot correspond to the frequencies of
% the signal X.
% The weights can be estimated in the same way as for the annihilating
% filter method or also using the equation (5.15) from the lecture notes.
% The second option is more stable to noise since we are "substracting" the
% noise component.
```

(b) The annihilating filter method can be used in the same way as for the previous case. The MUSIC method can be used as well. The only difference would be when estimating the weights because we cannot use straight forward the equation (5.15) from the lecture notes (try to see why).

Exercise 43. LINE SPECTRUM ESTIMATION: THE DUAL PROBLEM
Let $x(t)$ be a continuous periodic signal of period T,

$$x(t) = \sum_{n\in\mathbb{Z}} \sum_{k=0}^{M-1} a_k \delta(t - nT - t_k)$$

where $\delta(t)$ is a Dirac delta function. Assume that you want to use the annihilating filter method to estimate parameters $t_k,\ k = 0 \dots M-1$ from an appropriate set of the Fourier series coefficients.

(a) Compute the Fourier series coefficients $\widehat{x}[n]$ of $x(t)$.

(b) Write a system of equations that allows you to find t_k for $M = 3$. What is the minimum number of Fourier series coefficients required for a unique solution?

(c) How does the noisy case differ from the previous case, where the presence of noise was not considered?

Solution 43. SPECTRAL ESTIMATION

(a) The Fourier series coefficients of $x(t)$ are given by

$$\widehat{x}[m] = \frac{1}{T}\int_0^T \sum_{k=0}^{M-1} a_k \delta(t - t_k) e^{-j\frac{2\pi}{T}mt} dt = \frac{1}{T}\sum_{k=0}^{M-1} a_k e^{-j\frac{2\pi}{T}mt_k}$$

(b) For $M = 3$, we have

$$\widehat{x}[m] = \frac{1}{T}(a_0 e^{-j\frac{2\pi}{T}mt_0} + a_1 e^{-j\frac{2\pi}{T}mt_1} + a_2 e^{-j\frac{2\pi}{T}mt_2})$$

Using the annihilating filter method, we will choose a filter of length 4 being $[1, h_1, h_2, h_3]$. We have then $H(Z)X(m) = 0$. In matrix notation we get

$$\begin{bmatrix} \widehat{x}[3] & \widehat{x}[2] & \widehat{x}[1] & \widehat{x}[0] \\ \widehat{x}[4] & \widehat{x}[3] & \widehat{x}[2] & \widehat{x}[1] \\ \widehat{x}[5] & \widehat{x}[4] & \widehat{x}[3] & \widehat{x}[2] \end{bmatrix} \begin{bmatrix} 1 \\ h_1 \\ h_2 \\ h_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

We thus need at least 6 components of Fourier series coefficients to have a unique solution.

(c) As already discussed in class, the anhilating filter method is not robust to addition of noise. For small variance of the noise and large values of the a_k coefficients, the algorithm will still perform well, but as soon as the SNR becomes too small, the algorithm cannot be used anymore.

Exercise 44. Numerical Analysis of the Periodogram (Matlab)
Consider an AR(1) process $X[n] = cX[n-1] + W[n]$ where $c \leq 1$ (take $c = 0.9$, $X[0] \equiv 1$) and W to be an i.i.d. normally distributed noise of zero mean and unit variance.

(a) Write a function `pergram` that computes a periodogram $R_p(e^{j\omega})$ of a process of length M. Consider first the signal without noise $X[n] = cX[n-1]$. For $M = 256, 512, 1024$ plot in logarithmic scale R_p and compute the variance of R_p. Comment on your results. Repeat the experiment for the noisy signal $X[n] = cX[n-1] + W[n]$. Does the variance decrease as you increase M? Explain your answer.

(b) Compute the averaged periodogram of $N = 4$ segments of size $L = 256$. Plot the result, and compute its variance. Did the variance decrease with respect to the $M = 256$ case in the exercise 4.1.?

(c) Add a sinusoid $s[n] = A\sin(2\pi n/F)$ to the signal X, with $A = 5, F = 10$. Compute and plot the periodogram of the new signal for $M = 256, 512, 1024$. What do you notice? Compare the resulting component corresponding to the sinusoid with the component corresponding to the AR process. Repeat the experiment with the averaged periodogram. How do the two components mentioned above modify? Next, add one more sinusoid $\widetilde{s}[n]$ with amplitude A and with a frequency close to the frequency of $s[n]$ (F). Repeat the above experiment and comment on your results.

Solution 44. The periodogram

(a) The variance of the periodogram $R_p(e^{j\omega})$ does not decrease when the number of samples are increased. This can be directly observed from the fact that the fluctuations in the plots do not decrease as M increases from 256 to 1024. The variance of the periodogram $Var(R_p(e^{j\omega}))$ can be computed by considering several realizations of the process $X[n]$ and looking at the value of $R_p(e^{j\omega})$ for fixed ω.

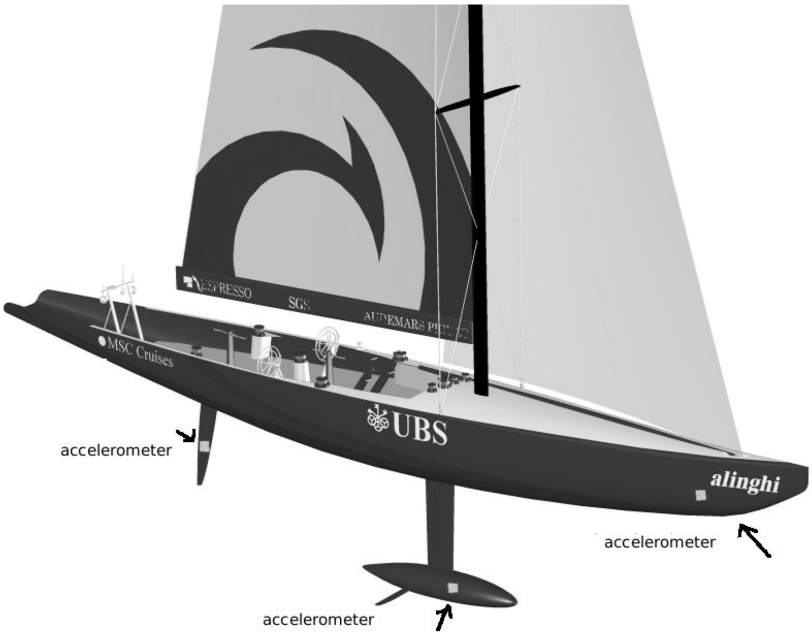
(b) The spectrum gets smoother by averaging and the variance decreases.

(c) The harmonic is well detected, while the spectrum of the AR signal is noisy. Averaging especially helps to smooth the component corresponding to the AR signal. When adding a second harmonic, if the resolution is not large enough, the two harmonics confound on the spectrum and are not detectable separately.

Exercise 45. ALINGHI I: VIBRATION CONTROL IN HIGH TECH SAILING BOATS (*20 points*)
High tech sailing boats, like Alinghi's one, are built using very sophisticated light materials that work in critical conditions, close to their break point. During the preliminary testing of the boat prototypes it is of foremost importance to accurately monitor the working conditions of such materials.

In view of the next America's cup challenge, one of the competitors is already performing sea testings of a new boat measuring the vibration of the hull and asked us to perform the data analysis.

We assume here that there are only three main vibrations in the boat hull and that we measure them using three accelerometers, as depicted in the figure below.



The vibration measured by the $k-$th accelerometer, $k = 1, 2, 3$, can be approximated by a complex exponential at frequency ω_k, *i.e.*, $e^{i\omega_k n}$.
We would like use the tools available for WSSprocesses. To do so, we have to model the accelerometer signals as a WSS.

1) By taking into account the uncertainty (randomness) on the origin of the complex exponential, precisely write the signals of the three accelerometers as WSSstochastic processes $V_k[n]$, $k = 1, 2, 3$.

In practice the signal of each accelerometer is not only composed of a single vibration but it also affected by the interferences of the vibrations measured by the two other accelerometers. Such interferences are often called "beats".
Consequently, the overall signal measured by the three accelerometers can be seen as the sum of the three complex exponential signals at frequencies ω_k, $k = 1, 2, 3$, plus the "beats" term which in our case is given by the sum of three complex exponentials at frequencies $\omega_2 - \omega_1$, $\omega_3 - \omega_1$, $\omega_3 - \omega_2$, as depicted in the diagram below.

Vibrations at ω_1, ω_2, ω_3 + Overall measured signal

"beats" term at $\omega_2 - \omega_1$, $\omega_3 - \omega_1$, $\omega_3 - \omega_2$

Notice that the beats term is due to the interferences between the four measured vibrations. Once again, we would like use the tools available for WSSprocesses and, to do so, we have to model the overall signal as a WSS.

2) By taking into account the uncertainty (randomness) on the origin of the complex exponential, precisely write the overall measured signal as WSSstochastic processes $X[n]$, and compute its mean.

After several measurements, the technical team have empirically established that

$$\frac{10}{100}\pi < \omega_1 < \frac{30}{100}\pi \qquad \frac{70}{100}\pi < \omega_2 < \frac{75}{100}\pi \qquad \frac{80}{100}\pi < \omega_3 < \pi$$

We are called to give a support to the technical team and our task is to provide a precise estimation for the vibration frequencies. In particular, we have to

- estimate the three vibration frequencies ω_k, $k = 1, \ldots, 3$;
- estimate the beats frequencies $\omega_2 - \omega_1$, $\omega_3 - \omega_1$, $\omega_3 - \omega_2$ in order to validate the beats model.

We start by using the simplest tool we know: the periodogram

3) How many samples N are necessary to perform the above required tasks?

In order to reduce the variance of the periodogram we adopt a Blackman-Turkey periodogram with a smoothing window. Such a smoothing window is non zero only over a support equal to $N/2$, where N is the number of samples considered.

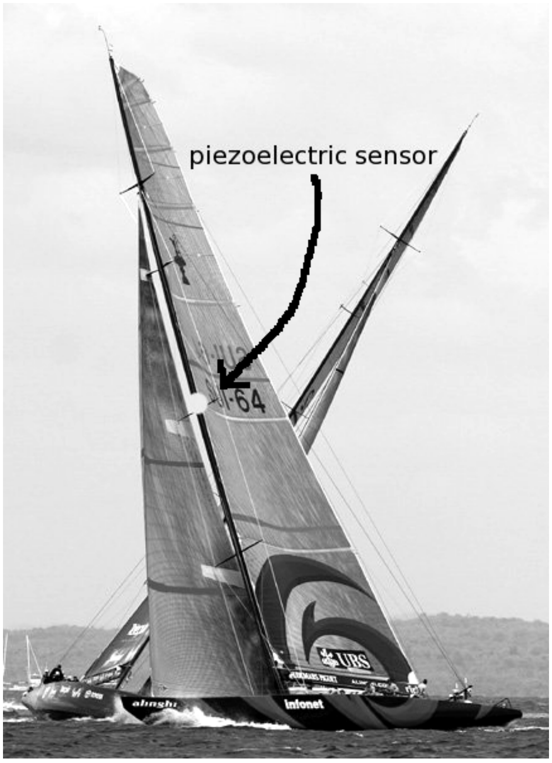
4) How many samples N do we need now to perform the above mentioned tasks?

5) What are the statistical properties of the Blackman-Turkey periodogram (variance, bias)?

Solution 45.

Exercise 46. ALINGHI II: MAST STRESS ANALYSIS (*16 points*)
The Mast (the large pole used to hold up the sails) is definitively a critical component of the sailing boat. Here again, the high technology materials used for its construction are pushed to their stress limit. During prototype testings, the behavior of the Mast must be monitored so to assure that it is properly dimensioned: if the mast breaks, the game is over (as for NZ team in the 2003 edition).

More precisely, we monitor the elongations of the Mast using a piezoelectric sensor positioned at its middle point, as depicted in the figure below.



Much of the information on the mast stress is contained in the power spectrum of the signal measured by the piezoelectric sensor. In particular such a power spectrum is smooth and can be approximated by a fractional polynomial

$$S(\omega) = \left.\frac{1}{C(z)}\right|_{z=e^{j\omega}}$$

1) Assuming that the approximation of a smooth spectrum (fractional polynomial) is correct, precisely describe a method for estimating the spectrum. More precisely we need to
   1.a) Estimate the number of parameters describing the spectrum (order of the polynomial, etc.)
   1.b) Estimate the value of such parameters
   1.c) Provide an estimation error

We then realize that the smooth spectrum (fractional polynomial) approximation is not exactly correct.

2) How this will affect the estimation of the number and values of the parameters?

The technical team complains that the method you have proposed is too complicated and ask you to use a periodogram based approach

3) Give precise arguments to defend your choice.

Solution 46.

(1) The mean is given by

$$H_0(z) = \sum 4(-\frac{1}{3})^k 2^{-k} = \frac{4}{1+\frac{1}{3}z^{-1}} = \frac{1}{P(z)} \Rightarrow Y_0(z) = \frac{1}{P(z)}X(z)$$

$X[n]$ is WSS, so $Y_0[n]$ is WSS too.

(2) From above it is clear that it is AR process.

(3) It has exactly the structure of correlation of AR processes; referee to the lecture notes.

(4) To check if the impulse response is really given by $h_0[k]$, one needs to build up the analysis filter and multiply it by the output of the system. The obtained signal should be white noise.

(5) The wiener filter is the optimal filter that could be designed. The process is WSS and such filter can be used.

Exercise 47. WHITE NOISE PERIODOGRAM
Let $y(t)$ be a zero-mean white noise with variance σ^2 and let

$$Y(\omega_k) = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} y(l) e^{-j\omega_k l}; \quad \omega_k = \frac{2\pi}{N} k \quad (k = 0, \dots, N-1)$$

denote its (normalized) DFT evaluated at the Fourier frequencies ω_k.

(a) Derive the covariances

$$E[Y(\omega_k) Y^*(\omega_r)], \quad k, r = 0, \dots, N-1$$

(b) Use the result of the previous calculation to conclude that the periodogram $\hat{\phi}(\omega_k) = |Y(\omega_k)|^2$ is an unbiased estimator of the PSD of $y(t)$.

(c) Explain whether the unbiasedness property holds for $\omega \neq \omega_k$ as well. Present an intuitive explanation of your finding.

Solution 47. WHITE NOISE PERIODOGRAM

(a)

$$\begin{aligned}
\mathbb{E}[Y(\omega_k) Y^*(\omega_r)] &= \mathbb{E}[\frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} y(l) e^{-j\omega_k l} \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} y^*(m) e^{j\omega_r m}] \\
&= \mathbb{E}[\frac{1}{N} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} y(l) y^*(m) e^{-j(\omega_k l - \omega_r m)}] \\
&= \frac{1}{N} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \mathbb{E}[y(l) y^*(m)] e^{-j(\omega_k l - \omega_r m)} \\
&= \frac{1}{N} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \sigma^2 \delta_{l-m} e^{-j(\omega_k l - \omega_r m)} \\
&= \frac{1}{N} \sigma^2 \sum_{l=0}^{N-1} e^{-jl \frac{2\pi}{N}(k-r)} \\
&= \frac{1}{N} \sigma^2 N \delta_{k-r} \\
&= \sigma^2 \delta_{k-r} \,.
\end{aligned}$$

(b) Since $y(t)$ is a zero-mean white noise with variance σ^2, we know that its PSD is equal to the variance:

$$S_y(\omega) = \sigma^2 \,.$$

To show that $\hat{\phi}(\omega_k)$ is an unbiased estimator of the PSD, we should show that its expectation is equal to the actual value of the PSD. We could do it in the following way:

$$\begin{aligned}
\mathbb{E}[\hat{\phi}(\omega_k)] &= \mathbb{E}[|Y(\omega_k)|^2] \\
&= \mathbb{E}[Y(\omega_k) Y^*(\omega_k)] \\
&= \sigma^2 \delta_{k-k} \\
&= \sigma^2 \,.
\end{aligned}$$

(c) Taking $\hat{\phi}(\omega) = |Y(\omega)|^2$ for any ω, we can find the expectation of the estimator $\hat{\phi}(\omega)$ in the following way:

$$\begin{aligned}
\mathbb{E}[\hat{\phi}(\omega)] &= \mathbb{E}[|Y(\omega)|^2] \\
\mathbb{E}[Y(\omega)Y^*(\omega)] &= \mathbb{E}[\frac{1}{\sqrt{N}}\sum_{l=0}^{N-1} y(l)e^{-j\omega l}\frac{1}{\sqrt{N}}\sum_{m=0}^{N-1} y^*(m)e^{j\omega m}] \\
&= \mathbb{E}[\frac{1}{N}\sum_{l=0}^{N-1}\sum_{m=0}^{N-1} y(l)y^*(m)e^{-j\omega(l-m)}] \\
&= \frac{1}{N}\sum_{l=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}[y(l)y^*(m)]e^{-j\omega(l-m)} \\
&= \frac{1}{N}\sum_{l=0}^{N-1}\sum_{m=0}^{N-1} \sigma^2\delta_{l-m}e^{-j\omega(l-m)} \\
&= \frac{1}{N}\sigma^2\sum_{l=0}^{N-1} e^{-j\omega 0} \\
&= \frac{1}{N}\sigma^2 N \\
&= \sigma^2 .
\end{aligned}$$

Since the expected value of the estimator $\hat{\phi}(\omega)$ is equal to the actual value of the PSD $S_y(\omega)$, it is unbiased for any frequency ω.

Exercise 48. Window selection for Blackman-Tukey method
Consider the case when the signal is composed of two harmonic components which are spaced in frequency by a distance larger than $1/N$. If you were to use a Blackman-Tukey method for spectral estimation, what window would you use if:

(a) The two spectral lines are closely-spaced in frequency, and they have similar magnitudes?

(b) The two spectral lines are not closely-spaced in frequency, and their magnitudes differ significantly?

Solution 48. Window selection for Blackman-Tukey method

(a) In order to discriminate between the two spectral lines of similar magnitudes spaced at a distance slightly larger than $1/N$, one needs to use a window which provides the best spectral resolution, that is, the rectangular window (or the unmodified periodogram).

(b) Although using a window with high spectral resolution is beneficial, it has its downsides. Namely, such windows have worse side-lobe attenuation, which in the case of Blackman-Tukey method means that the spectral lines leak more energy to the surrounding frequencies.

In the problem at hand, if the used window has low side-band attenuation, the spectral line with significantly higher magnitude can leak enough energy to the frequencies around the weaker spectral line as to make the weaker spectral line less pronounced and harder to detect. In order to reduce the contribution of the more powerful spectral line to the far-away frequencies, one should use windows which have a better side-lobe attenuation, such as Hamming, von Hann, variants of Kaiser window etc.

Exercise 49. SPECTRAL FACTORIZATION AND ESTIMATION
Let $\{X[n]\}_{n\in\mathbb{Z}}$ be a centered AR process with power spectral density of the form

$$S_X(\omega) = \frac{b}{(1+a_1^2-2a_1\cos\omega)(1+a_2^2-2a_2\cos\omega)}, \quad |a_1|<1,\ |a_2|<1,\ b>0,$$

where a_1, a_2 and b are unknown real-valued parameters.

(a) Give the canonical representation of the process $X[n]$

$$P(z)X[n] = W[n]$$

Give the whitening filter $P(z)$ and the variance σ^2 of the noise process $\{W[n]\}_{n\in\mathbb{Z}}$.

(b) Give the procedure to determine the parameters a_1, a_2 and b of the AR process $X[n]$, and to estimate the power spectral density $S_X(\omega)$.

Solution 49. SPECTRAL FACTORIZATION AND ESTIMATION
The power spectral density $S_X(\omega)$ can be transformed in the following way:

$$\begin{aligned} S_X(\omega) &= \frac{b}{(1+a_1^2-2a_1\cos\omega)(1+a_2^2-2a_2\cos\omega)} \\ &= \frac{b}{|1-a_1e^{-j\omega}|^2|1-a_2e^{-j\omega}|^2} \\ &= \frac{b}{|(1-a_1e^{-j\omega})(1-a_2e^{-j\omega})|^2}. \end{aligned}$$

(a) The power spectral density of an AR process has the form

$$S(\omega) = \frac{1}{|P(e^{j\omega})|^2}\sigma_W^2$$

where $P(z)$ is the minimum phase whitening filter, and σ_W^2 is the noise variance. Since $|a_1|<1$ and $|a_2|<1$, the polynomial $P(z) = (1-a_1z^{-1})(1-a_2z^{-1})$ is strictly minimum phase, and since $b>0$, we can see that the PSD $S_X(\omega)$ corresponds to an AR process $P(z)X[n] = W[n]$, whose whitening filter is given by

$$P(z) = 1-(a_1+a_2)z^{-1}+a_1a_2z^{-2},$$

with the noise $W[n]$ having the variance $\sigma_W^2 = b$.

(b) Making the substitutions $p_1 = a_1+a_2$ and $p_2 = -a_1a_2$, we can write $P(z) = 1-p_1z^{-1}-p_2z^{-2}$. The parameters p_1, p_2 and b can be determined by solving the following Yule-Walker equations:

$$\begin{aligned} b + p_1\hat{R}_X[1] + p_2\hat{R}_X[2] &= \hat{R}_X[0] \\ p_1\hat{R}_X[0] + p_2\hat{R}_X[1] &= \hat{R}_X[1] \\ p_1\hat{R}_X[1] + p_2\hat{R}_X[0] &= \hat{R}_X[2], \end{aligned}$$

where $\hat{R}_X[0]$, $\hat{R}_X[1]$ and $\hat{R}_X[1]$ are the empirical correlation estimates at lags 0, 1 and 2.

Once the parameters p_1, p_2 and b have been determined, the parameters a_1 and a_2 can be determined by solving the non-linear system

$$\begin{aligned} a_1 + a_2 &= p_1 \\ a_1 a_2 &= -p_2 \,, \end{aligned}$$

under the constraints that $|a_1| < 1$ and $|a_2| < 1$. Furthermore, the estimated power spectral density $S_X(\omega)$ has the form

$$S_X(\omega) = \frac{b}{|1 - p_1 e^{-j\omega} - p_2 e^{-2j\omega}|^2} \,.$$

Chapter 6

Transforms

Exercise 50. Discrete Cosine Transform in Matlab
The KLT is a signal-dependent transform. This property is inconvenient if a signal has to be transmitted because the receiver needs to know both the transform coefficients and the transform basis vectors. The Discrete Cosine Transform (DCT) is signal-independent and very close to the optimal KLT in terms of correlation of the transform coefficients.

(a) For the jointly Gaussian sequence of vectors derived in Exercise 4 of the numerical part, calculate the DCT coefficients. *Hint:* Use the Matlab function `dct`.

(b) Evaluate the correlation matrix of the DCT coefficients. How far is it from the KLT correlation matrix?

Solution 50. Discrete Cosine Transform

(a) The Matlab function `dct` performs the DCT along columns of the input matrix, thus

```
z1 = dct(y);
```

(b) The correlation is calculated similarly as in Exercise 4:

```
Rz1 = (z1*z1')/M;
```

Notice that the correlation matrix of the DCT coefficients is not diagonal, but the diagonal terms carry most of energy in the transform domain. Comparing the DCT coefficient correlation to the KLT coefficient correlation produced in Exercise 4, we can conclude that the DCT coefficients are more correlated. However, the main advantage of the DCT is fixed structure, that is, the basis vectors do not depend on the signal.

Exercise 51. CORRELATING AND DECORRELATING SIGNALS
In this exercise, we will see that a signal can be both correlated and decorrelated by applying a suitable linear transform, where in the later case the optimal transform is the Karhunen–Loeve transform (KLT). KLT in the literature of signal processing is basically equivalent to PCA.

(a) Consider an i.i.d. (independent, identically distributed) sequence of random variables $X_0, X_1, \ldots, X_{N-1}$ ($\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = 1$, for $i = 0, 1, \ldots, N-1$). Define a new set of random variables $\mathbf{Y} = [Y_0, Y_1, \ldots, Y_{N-1}]^T$ as

$$\mathbf{Y} = \mathbf{A} \cdot \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix},$$

where

$$\mathbf{A} = \begin{bmatrix} \alpha_{0,0} & \alpha_{0,1} & \cdots & \alpha_{0,N-1} \\ \alpha_{1,0} & \alpha_{1,1} & \cdots & \alpha_{1,N-1} \\ \vdots & \vdots & & \vdots \\ \alpha_{N-1,0} & \alpha_{N-1,1} & \cdots & \alpha_{N-1,N-1} \end{bmatrix}$$

is a real square matrix.

Show that the correlation function satisfies:

$$R_{i,j} = E[Y_i \cdot Y_j] = \sum_{k=0}^{N-1} \alpha_{i,k} \cdot \alpha_{j,k},$$

for $i, j = 0, 1, \ldots, N-1$.

(b) Show that the following equality holds:

$$\det(\mathbf{A}) = \prod_{i=0}^{N-1} \lambda_i^{1/2},$$

where λ_is are eigenvalues of the correlation matrix $\mathbf{R}_y$.

(c) Consider a time sequence of random vectors $\mathbf{Y}[n] = [Y_0[n], Y_1[n], \ldots, Y_{N-1}[n]]^T$. The KLT of the random signal $\mathbf{Y}[n]$ is obtained as $\mathbf{Z}[n] = \mathbf{T} \cdot \mathbf{Y}[n]$, where the rows of the matrix $\mathbf{T}$ are the eigenvectors of the correlation matrix of the signal $\mathbf{Y}[n]$ (sorted in descending order of the corresponding eigenvalues).

Show that the resulting vector coefficients $\mathbf{Z}[n]$ are uncorrelated. Are they independent?

Solution 51. CORRELATING AND DECORRELATING SIGNALS

(a) The set of random variables $\mathbf{Y}$ is determined as

$$Y_i = \sum_{k=0}^{N-1} \alpha_{i,k} \cdot X[k].$$

The correlation function is given by

$$R_{i,j} = \mathbb{E}[Y_i \cdot Y_j] = \mathbb{E}[\sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \alpha_{i,k}\alpha_{j,l} X_k X_l]$$
$$= \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \alpha_{i,k}\alpha_{j,l} \mathbb{E}[X_k X_l].$$

Since the random variables X_i are normalized independent Gaussian variables, we know that $\mathbb{E}[X_k] = 0$ and $\mathbb{E}[X_k^2] = 1$. It follows that $\mathbb{E}[X_k X_l] = 0$ for $k \neq l$. Therefore, we can write

$$\mathbb{E}[X_k X_l] = \delta[k-l].$$

The correlation is now given by

$$R_{i,j} = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \alpha_{i,k}\alpha_{j,l}\delta[k-l],$$

or, equivalently

$$R_{i,j} = \sum_{k=0}^{N-1} \alpha_{i,k}\alpha_{j,k}.$$

(b) From (a) we can see that the correlation matrix $\mathbf{R}_Y$ is defined as $\mathbf{R}_Y = \mathbf{A} \cdot \mathbf{A}^T$. If we apply the det operator, we obtain

$$\det(\mathbf{R}_Y) = \det(\mathbf{A}) \cdot \det\left(\mathbf{A}^T\right) = (\det(\mathbf{A}))^2. \tag{6.1}$$

The correlation matrix $\mathbf{R}_y$ can also be expressed in terms of its eigenvalues and eigenvectors as

$$\mathbf{R}_Y = \mathbf{V}_Y \cdot \mathbf{\Lambda}_Y \cdot \mathbf{V}_Y^T,$$

where $\mathbf{V}_y$ is a matrix, which contains the eigenvectors as columns and $\mathbf{\Lambda}_y$ is a diagonal matrix with the eigenvalues along the diagonal.

Similarly, we can write

$$\det(\mathbf{R}_Y) = \det(\mathbf{V}_Y) \cdot \det(\mathbf{\Lambda}_Y) \cdot \det\left(\mathbf{V}_Y^T\right) = \prod_{i=0}^{N-1} \lambda_i \tag{6.2}$$

because $\mathbf{\Lambda}_Y$ is diagonal and $\det(\mathbf{V}_Y) = 1$. Therefore from (6.1) and (6.2), we have

$$\det(\mathbf{R}_Y) = (\det(\mathbf{A}))^2 = \prod_{i=0}^{N-1} \lambda_i.$$

It follows that

$$\det(\mathbf{A}) = \prod_{i=0}^{N-1} \lambda_i^{1/2}.$$

(c) The KLT matrix $\mathbf{T}$ is given by $\mathbf{T} = \mathbf{V}_Y^T$ because it contains the eigenvectors of the correlation matrix as the rows. Therefore, we can write

$$\mathbf{Z}[n] = \mathbf{V}_Y^T \cdot \mathbf{Y}[n].$$

The correlation $\mathbf{R}_Z$ is given by $\mathbf{R}_Z = \mathbb{E}\left[\mathbf{Z}[n] \cdot \mathbf{Z}^T[n]\right]$. It follows

$$\mathbf{R}_Z = \mathbb{E}\left[\mathbf{V}_Y^T \cdot \mathbf{Y}[n] \cdot \mathbf{Y}^T[n] \cdot \mathbf{V}_Y\right] = \mathbf{V}_Y^T \cdot \mathbf{R}_Y \cdot \mathbf{V}_Y.$$

From (b) we know that $\mathbf{R}_Y = \mathbf{V}_Y \cdot \mathbf{\Lambda}_Y \cdot \mathbf{V}_Y^T$. Therefore

$$\mathbf{R}_Z = \mathbf{V}_Y^T \cdot \mathbf{V}_Y \cdot \mathbf{\Lambda}_Y \cdot \mathbf{V}_Y^T \cdot \mathbf{V}_Y = \mathbf{\Lambda}_Y,$$

because $\mathbf{V}_Y^T \cdot \mathbf{V}_Y = \mathbf{I}$. The correlation matrix $\mathbf{R}_Z$ is diagonal and, thus, the variables $Z_i[n]$, for $i = 0, 1, \ldots, N-1$, are uncorrelated. Since the random variables are Gaussian, uncorrelation is equivalent to independence.

Exercise 52. USING THE KARHUNEN-LOÈVE TRANSFORM IN MATLAB

(a) Generate an i.i.d. sequence of 5 normalized Gaussian random variables X_0, X_1, X_2, X_3 and X_4.

(b) Using the sequence generated in (a) and the results from the theoretical part, generate a sequence $\mathbf{Y}[n]$ of $M = 10000$ i.i.d. jointly Gaussian random vectors of size $N = 5$ (the corresponding signal matrix has the size $N \times M$) with the following correlation matrix:

$$\mathbf{R}_Y = \begin{bmatrix} 1.9 & 0.5 & 0.3 & 0.2 & 0.05 \\ 0.5 & 2.3 & 0.4 & 0.2 & 0.1 \\ 0.3 & 0.4 & 1.5 & 0.9 & 0.7 \\ 0.2 & 0.2 & 0.9 & 1.1 & 0.8 \\ 0.05 & 0.1 & 0.7 & 0.8 & 1.2 \end{bmatrix}.$$

Hint: You may use the Matlab function `eig` to calculate the eigenvectors and eigenvalues.

(c) Evaluate the correlation matrix $\hat{\mathbf{R}}_Y$ of the generated sequence. How far is it from the specified $\mathbf{R}_Y$? Compute $\hat{\mathbf{R}}_y$ for different values of M and compare it to $\mathbf{R}_Y$.

(d) Derive the KLT matrix $\mathbf{T}$ based on the evaluated correlation $\hat{\mathbf{R}}_Y$. Calculate the transform coefficients $\mathbf{Z} = \mathbf{T} \cdot \mathbf{Y}$. Now, evaluate the resulting correlation of the transform coefficients $\hat{\mathbf{R}}_Z$. Is it diagonal?

Solution 52. USING THE KARHUNEN-LOÈVE TRANSFORM IN MATLAB

(a) In Matlab:

```
N = 5; x = randn(N,1);
```

(b) First, let us generate the Gaussian normalized sequence $\mathbf{X}[n]$ of the length M:

```
M = 10000; x = randn(N,M);
```

Now, choose the matrix $\mathbf{A}$, as $\mathbf{A} = \mathbf{V}_Y \cdot \mathbf{\Lambda}_Y^{1/2}$, where $\mathbf{V}_Y$ and $\mathbf{\Lambda}_Y$ are eigenvectors and eigenvalues of the autocorrelation matrix $\mathbf{R}_Y$. This ensures that the correlation of the variables Y_is is given by $\mathbf{R}_Y$ in limit (Exercise 1).

```
Ry = [1.9 0.5 0.3 0.2 0.05;
      0.5 2.3 0.4 0.2 0.1;
      0.3 0.4 1.5 0.9 0.7;
      0.2 0.2 0.9 1.1 0.8;
      0.05 0.1 0.7 0.8 1.2];
[Vy,Ly] = eig(Ry); A = Vy*Ly^0.5; y = A*x;
```

(c) The correlation of the generated sequence $\mathbf{y}$ is evaluated by

```
Ry1 = (y*y')/M;
```

The correlation $\hat{\mathbf{R}}_y$ approximates the expected correlation $\mathbf{R}_y$. They are not exactly the same because of the finite length of the Gaussian sequence $\mathbf{x}$. As the length M grows, the approximated correlation $\hat{\mathbf{R}}_y$ is closer to the expected correlation $\mathbf{R}_y$.

(d) The KLT matrix $\mathbf{T}$ contains the eigenvectors of the estimated correlation matrix $\hat{\mathbf{R}}_y$ as rows in descending order of the corresponding eigenvalues. The KLT and the correlation $\mathbf{R}_z$ are calculated by

```
% Compute the eigenvectors and eigenvalues of the estimated correlation matrix Ry1
[Vy1,Ly1] = eig(Ry1);
% Then, sort the eigenvalues
[Lsorted,I] = sort(diag(Ly1));
% Arrange them in descending order (sort gives ascending order)
I = I(length(I):-1:1);
% Take the corresponding columns from Vy1 and put them as rows in T
T = Vy1(1:N,I)';
% Apply the KLT
z = T*y;
% Compute the correlation Rz
Rz = (z*z')/M;
```

The correlation $\mathbf{R}_z$ is diagonal. This is expected since we used the estimated correlation matrix $\hat{\mathbf{R}}_y$. The KLT is obviously signal-dependent, because it is constructed using the properties of the generated signal.

Exercise 53. KARHUNEN-LOÉVE TRANSFORM
Consider a block of Gauss-Markov first-order random variables of size 4: $\mathbf{X} = [X_0, X_1, X_2, X_3]^T$. Its correlation matrix is given by:

$$R_{\mathbf{X}} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

where ρ is the correlation coefficient between two adjacent random variables.
Now take two sub-transforms of size 2×2, namely, KLT of $[X_0, X_1]^T$ and KLT of $[X_2, X_3]^T$ to produce $\mathbf{Y} = [Y_0, Y_1, Y_2, Y_3]^T$.

(a) What is the resulting transform?

(b) Calculate the resulting correlation matrix $R_{\mathbf{Y}}$.

(c) Calculate the coding gain associated to $R_{\mathbf{Y}}$, i.e. the two sub-transforms, and compare with the coding gain of the KLT.

Solution 53. KARHUNEN-LOÉVE TRANSFORM

(a) The autocorrelation of $[X_0, X_1]$ is

$$R_{[X_0,X_1]} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

The eigenvectors of this 2×2 symmetric matrix are always $\mathbf{v}_0 = 1/\sqrt{2}[1,1]^T$ and $\mathbf{v}_1 = 1/\sqrt{2}[1,-1]^T$ Using

$$T = 1/\sqrt{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

as a transform, we get

$$R_{[Y_0,Y_1]} = \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}$$

The resulting transform to be used is

$$\begin{bmatrix} Y_0 \\ Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = 1/\sqrt{2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

(b) The resulting $R_{\mathbf{Y}}$ is obtained by

$$R_{\mathbf{Y}} = TR_{\mathbf{X}}T^T.$$

Calculations give

$$R_{\mathbf{Y}} = \begin{bmatrix} 1+\rho & 0 & \rho/2(1+\rho)^2 & \rho/2(1-\rho^2) \\ 0 & 1-\rho & \rho/2(\rho^2-1) & -\rho/2(1-\rho)^2 \\ \rho/2(1+\rho)^2 & \rho/2(\rho^2-1) & 1+\rho & 0 \\ \rho/2(1-\rho^2) & -\rho/2(1-\rho)^2 & 0 & 1-\rho \end{bmatrix}$$

(c)

$$G_Y = \frac{1}{\left(\Pi_{i=0}^{3} R_Y^2(i)\right)^{1/4}} = \frac{1}{\left((1+\rho)^2(1-\rho)^2\right)^{1/4}}$$

$$G_{KTL} = \frac{1}{\left(det\ R_x\right)^{1/4}} = \frac{1}{\left(1-3\rho^2+3\rho^4-\rho^6\right)^{1/4}}$$

Remark that $G_{KTL} > G_Y$ for any ρ. This is in agreement with the fact that the KLT maximizes the coding gain.

Exercise 54. KLT OF CIRCULANT CORRELATION MATRICES
Let X be a real periodic sequence of period $N = 4$ with correlation matrix R_x:

$$R_x = \begin{bmatrix} 1 & 0.4 & 0.2 & 0.4 \\ 0.4 & 1 & 0.4 & 0.2 \\ 0.2 & 0.4 & 1 & 0.4 \\ 0.4 & 0.2 & 0.4 & 1 \end{bmatrix}$$

(a) Compute its KLT, that is, the transform T that diagonalizes R_x.

(b) Consider now the DFT matrix S_N of size $N = 4$. Compute $S_N^* R_x S_N$. What do you obtain? Recall that the DFT can be formulated as a complex matrix multiplication $X[k] = S_N x[n]$ where the DFT matrix S_N is given by $S_N[k,n] = W_N^{-kn}$.

(c) Compare both solutions. What can you conclude?

Solution 54. KLT OF CIRCULANT CORRELATION MATRICES

(a) The KLT matrix T is given by the eigenvectors of R_x:

$$T = \begin{bmatrix} -1/2 & 1/2 & -1/2 & 1/2 \\ 0 & -\sqrt{(2)}/2 & 0 & \sqrt{(2)}/2 \\ \sqrt{(2)}/2 & 0 & -\sqrt{(2)}/2 & 0 \\ 1/2 & 1/2 & 1/2 & 1/2 \end{bmatrix}$$

To show that this is indeed a KLT matrix we compute TR_xT^T:

$$TR_xT^T = \begin{bmatrix} 0.4 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix},$$

which is a diagonal matrix.

(b) The DFT matrix $S_N[k,n] = W_N^{-kn}$ of size $N = 4$ is given by:

$$D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix}.$$

If we compute now $S_N^* R_x S_N$ we obtain:

$$S_N^* R_x S_N = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 3.2 & 0 & 0 \\ 0 & 0 & 1.6 & 0 \\ 0 & 0 & 0 & 3.2 \end{bmatrix},$$

which is also a diagonal matrix.

(c) Both transforms T and S_N give a diagonal correlation matrix and can be used as a decorrelation transform. However, the DFT matrix is constant for a given N and much easier to compute that the KLT matrix. However, the DFT matrix does not always produce the same results of the KLT. This exercise is a particular case where X is periodic and R_x is a circulant matrix. The reason is that the DFT matrix diagonalizes ANY circulant matrix. Therefore, if R_x is a circulant matrix, the DFT matrix is preferable as a decorrelation transform.

Exercise 55. KARHUNEN-LOÉVE TRANSFORM
Let R_x be the correlation matrix of a real periodic sequence $\mathbf{x}$ of period N.

(a) Calculate R_x for the case when $N = 2$ and $\mathbf{R_x} = [1, 0.5]$. Find the KLT of $\mathbf{x}$, that is, $\mathbf{y} = T\mathbf{x}$.

(b) Assume next that $N = 4$ and the correlation matrix R_x of the sequence $\mathbf{x}$ is given by

$$R_x = \begin{bmatrix} 1 & 0.4 & 0.2 & 0.4 \\ 0.4 & 1 & 0.4 & 0.2 \\ 0.2 & 0.4 & 1 & 0.4 \\ 0.4 & 0.2 & 0.4 & 1 \end{bmatrix} \tag{6.3}$$

Give a 4×4 matrix T that diagonalizes R_x. What is the resulting correlation matrix R_y?

Hint: Note that the DFT matrix diagonalizes any circulant matrix.

Solution 55. KARHUNEN-LOÉVE TRANSFORM
Since $x[n]$ is periodic, the correlation function is also periodic, i.e., $R_x[n] = R_x[n+kN]$, $\forall n, k \in \mathbb{Z}$. Therefore, if we use blocks of N consecutive samples of $x[n]$, we obtain vectors whose correlation matrix is circulant.

(a) In this case $N = 2$ and $R_x[n] = [1, 0.5]$ and the correlation matrix is

$$\mathbf{R_x} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix},$$

and the KLT is the matrix

$$\mathbf{H} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Remark that this is also the DFT of size 2 (properly scaled).

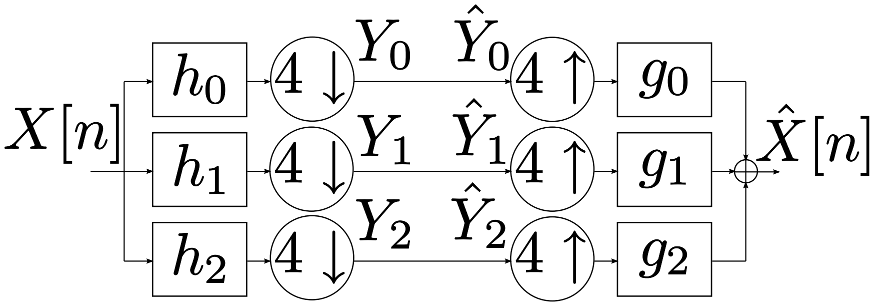
(b) In this case, we should use the DFT of size 4 and normalize $\sqrt{4}$, i.e.,

$$\mathbf{H} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix}$$

we have that $\mathbf{Y} = \mathbf{H}^{\star}\mathbf{X}$; therefore,

$$\mathbf{R_Y} = \mathbf{H}^{\star}\mathbf{R_x}\mathbf{H} = \begin{bmatrix} 2 & & & \\ & 0.8 & & \\ & & 0.4 & \\ & & & 0.8 \end{bmatrix}$$

Exercise 56. KLT Filter Bank
Consider the system:



where $X[n]$ is a zero-mean, Gaussian, stationary process with correlation function

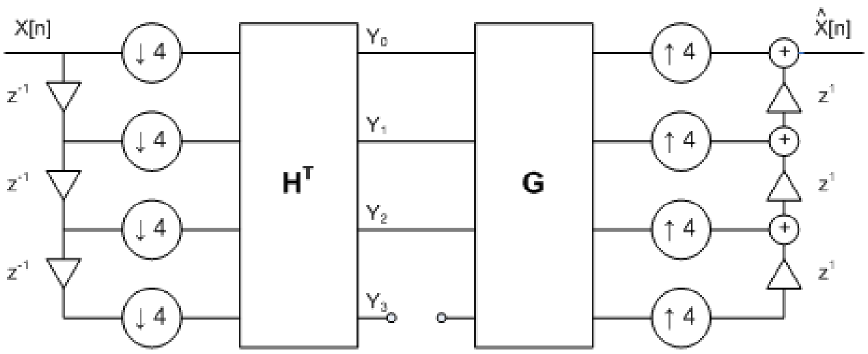
$$R_X[n] = \begin{cases} 3 & \text{if } n = 0 \\ 2 & \text{if } n = \pm 1 \\ 1 & \text{if } n = \pm 2 \\ 2 & \text{if } n = \pm 3 \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine a set of filters h_0, h_1, h_2, g_0, g_1, g_2 in order to minimize the average expectation of the reconstruction error $E[n] = X[n] - \hat{X}[n]$. (For simplicity, take the h_i's causal and the g_i's anti-causal. A solution with real filter coefficients exists but a complex solution is also accepted.)

(b) What is the reconstruction error for the filters in your answer to the previous question? What is the probability that the reconstructed process $\hat{X}$ is equal to the input process X? Is the process $\hat{X}$ stationary?

(c) Suppose that we decide to code the processes Y_0, Y_1, Y_2 with a finite number of bits. We assume that we use an ideal set of coders to compute the coded quantities $\hat{Y}_i$ from Y_i for $i = 0, 1, 2$. If we call $E_i = Y_i - \hat{Y}_i$ the coding error, we assume that the error variance is $\sigma_{E_i}^2 = \sigma_{Y_i}^2 2^{-2r_i}$, where r_i is the rate in bits per sample assigned to the coder i. We impose the rates r_i so that $\sigma_{E_0}^2 = \sigma_{E_1}^2 = \sigma_{E_2}^2$. We want to compare the advantage of using the transformation to directly code the process X. Assume that we use an average rate r to code the components Y_i, i.e. $4r = r_0 + r_1 + r_2$. What is the average reconstruction error variance, $\mathbb{E}[E[n]^2]$ as a function of the average rate r? Compare with the case where you code directly $X[n]$ using the same rate r (assume similarly that $\sigma_E^2 = \sigma_X^2 2^{-2r}$). Which of the two solutions would you prefer?

Solution 56. KLT Filter Bank
We remark that the given diagram is equivalent to a block transform as depicted in the following figure.



Remark that it is the down-sampling factor (i.e., 4) that determines the size of the transformation. Since the filter bank has only 3 branches, one of the transformed coefficients (Y_3 in the figure) is discarded and replaced with zero at the synthesis.

(a) The coefficients of the filter h and g are in the columns of the matrix H, i.e.,

$$H_i(z) = \sum_{j=0}^{3} h_{ji} z^{-j} \qquad G_i(z) = \sum_{j=0}^{3} h_{ji} z^{j},$$

so we have to determine $\mathbf{H}$ in order to minimize the average of the expected error. We know that the solution is the KLT. This is determined by diagonalizing the correlation matrix:

$$\mathbf{R_x} = \begin{bmatrix} 3 & 2 & 1 & 2 \\ 2 & 3 & 2 & 1 \\ 1 & 2 & 3 & 2 \\ 2 & 1 & 2 & 3 \end{bmatrix}.$$

This is a circulant matrix and we know it is diagonalized by the Fourier matrix:

$$\mathbf{F} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix} = \frac{1}{2} \begin{bmatrix} f_0 & f_1 & f_2 & f_3 \end{bmatrix}.$$

The eigenvalues can be computed by calculating:

$$\mathbf{F}^{\star}\mathbf{R_x}\mathbf{F} = \Lambda = \begin{bmatrix} 8 & & & \\ & 2 & & \\ & & 0 & \\ & & & 2 \end{bmatrix}.$$

To determine the KLT, we have to sort the eigenvectors according to a decreasing order of the eigenvalues and we have to normalize them to have unit norm. This gives:

$$\mathbf{H} = \frac{1}{2} \begin{bmatrix} f_0 & f_1 & f_3 & f_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{j}{2} & -\frac{j}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{j}{2} & \frac{j}{2} & -\frac{1}{2} \end{bmatrix},$$

and the first three columns give the filter coefficients. This solution gives complex coefficients for the filter H_1 and H_2.

(b) For this particular case, the eigenvalue corresponding to the removed component is zero; therefore, the expected error on the block error norm is:

$$\mathbb{E}\left[\left\|\overline{E}[n]\right\|^2\right] = h_3^T \mathbf{R_x} h_3 = 0,$$

and the reconstruction error is zero. This is due to the particular choice of the correlation function of the input process, which corresponds to a redundancy of the samples.

The probability of perfect reconstruction is:

$$P\left\{\hat{x}[n] = x[n]\right\} = P\left\{E[n] = 0\right\} = 1.$$

Remark that this probability would be zero if the error variance was larger than zero. Since the input process is perfectly reconstructed, the output process is stationary as the input process.

(c) We compute the variances of the coefficients Y_0, Y_1, Y_2. These are the first three diagonal terms of $\mathbf{H}^\star \mathbf{R_x} \mathbf{H}$, i.e., $\sigma_{Y_0}^2 = 8$, $\sigma_{Y_1}^2 = 2$, $\sigma_{Y_2}^2 = 2$. We have, computing the error variances $\sigma_{E_0}^2$, $\sigma_{E_1}^2$, $\sigma_{E_2}^2$.

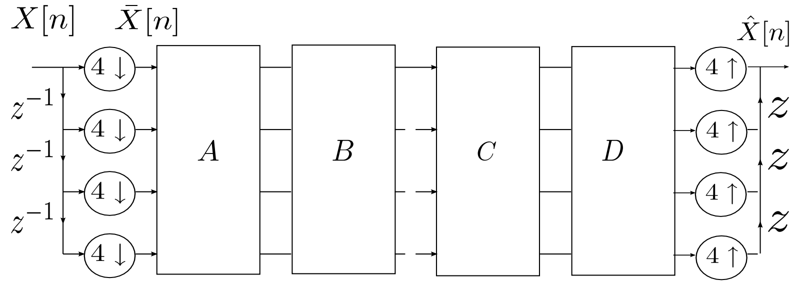
$$8 \cdot 2^{-2r_0} = 2 \cdot 2^{-2r_1} = 2 \cdot 2^{-2r_2};$$

therefore, $r_1 = r_2$ and $r_0 = r_1 + 1$. The average rate r is $r = \frac{r_0+r_1+r_2}{4} = \frac{3r_1+1}{4}$ and $r_1 = \frac{4r-1}{3}$. The average reconstruction error variance d is:

$$d = \frac{\sigma_{E_0}^2+\sigma_{E_1}^2+\sigma_{E_2}^2}{4} = \frac{3}{4}\sigma_{E_1}^2 = \frac{3}{2} 2 \cdot 2^{-2r_1} = \frac{3}{2} 2 \cdot 2^{-2\frac{4r-1}{3}} = 3\sqrt[3]{4} 2^{-\frac{8}{3}r}.$$

If $x[n]$ is coded directly, the error variance is $d_x = \sigma_x^2 2^{-2r} = 3 \cdot 2^{-2r}$ and the coding gain is, in this case, $G = \frac{d_x}{d} = \frac{3 \cdot 2^{-2r}}{3\sqrt[3]{4} 2^{-\frac{8}{3}r}} = 2^{\frac{2}{3}(r-1)}$, which is always larger than one (remark that $r > 1$, for the condition on r_0). This means that it is always convenient to use the transformation instead of coding directly the input process.

Exercise 57. KLT FILTER BANK



In the diagram $X[n]$ is a Gaussian stationary process with zero mean and correlation

$$R_X[n] = \frac{1}{2^{|n|}} + \frac{3}{8}\delta[n-3] + \frac{3}{8}\delta[n+3]$$

The blocks A, B, C and D transform the input vector to the output vector by matrix multiplication. After block B, three entries of the resultant vector are set to zero before entering in block C. The matrix B and C are given by

$$B = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & -1 & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & -1 & 0 & 0 \end{bmatrix}$$

Answer the following questions.

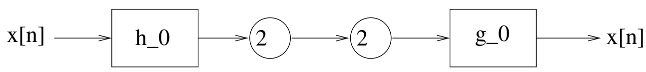
(a) Is $\hat{X}[n]$ a Gaussian process? Is it stationary?

(b) Compute the correlation matrix of the random vector $\bar{\boldsymbol{X}}[n]$, say $\tilde{R}_X[n]$ Can you diagonalize it (find the eigenvalue decomposition) without using Matlab?

(c) Find two possible matrices A and D, such that $\hat{X}[n]$ approaches $X[n]$ with minimum error variance, i.e. $\mathbb{E}\left[(X[n] - \hat{X}[n])^2\right]$ should be minimal.

(d) Find a geometric interpretation of the transformations BA and DC.

(e) Consider the following diagram



Find the filters H_0 and G_0 such that the system is equivalent to the one in the previous figure.

Solution 57. KLT FILTER BANK

(a) $\hat{X}[n]$ is a Gaussian process. Note that the outputs of the transform D are jointly Gaussian since they are linear combinations of the input process $X[n]$ (more precisely $X[n]$, $X[n-1]$, $X[n-2]$, $X[n-3]$). Hence the output of the parallel to serial converter is a Gaussian process. However $\hat{X}[n]$ is not stationary in general, since the reconstruction is not perfect and the upsamplers in the last stage break the stationarity.

(b) Observe that

$$\hat{\boldsymbol{X}}[n] = \begin{bmatrix} X[n] \\ X[n-1] \\ X[n-2] \\ X[n-3] \end{bmatrix}$$

and

$$R_X[0] = 1 \qquad R_X[1] = \frac{1}{2} \qquad R_X[2] = \frac{1}{4} \qquad R_X[3] = \frac{1}{2}. \tag{6.4}$$

Hence

$$\tilde{R}_X = \begin{bmatrix} 1 & 1/2 & 1/4 & 1/2 \\ 1/2 & 1 & 1/2 & 1/4 \\ 1/4 & 1/2 & 1 & 1/2 \\ 1/2 & 1/4 & 1/2 & 1 \end{bmatrix}$$

Observe that the correlation matrix $\tilde{R}_X$ is circulant and we know how to find the eigenvalues and eigenvectors of circulant matrices. The normalized eigenvectors are given by

$$w_k = \frac{1}{2} \begin{bmatrix} 1 \\ e^{j\frac{\pi k}{2}} \\ e^{j\pi k} \\ e^{j\frac{3\pi k}{2}} \end{bmatrix} \qquad k = 0, 1, 2, 3$$

and the corresponding eigenvalues are given by

$$\lambda_k = \sum_{i=0}^{3} \tilde{R}_X[i] e^{-j\frac{2\pi k}{4} i} \qquad k = 0, 1, 2, 3$$

thus

$$\lambda_0 = 9/4 \qquad \lambda_1 = 3/4 \qquad \lambda_2 = 1/4 \qquad \lambda_3 = 3/4.$$

Hence we have the decomposition

$$\tilde{R}_X = F \Lambda F^*$$

where $F = [w_0 \;\; w_1 \;\; w_2 \;\; w_3]$ is the 4×4 Fourier matrix and $\Lambda = \text{diag}(\lambda_0, \lambda_1, \lambda_2, \lambda_3)$.

(c) From the discussion in class, we know that we have the following conditions on the transforms BA and CD in order for the error variance $\mathbb{E}\left[(X[n] - \hat{X}[n])^2\right]$ to be minimum,

$$((BA)^T)_0 = w_0 \qquad (DC)_0 = w_0$$

where $((BA)^T)_0$ and $(DC)_0$ denote the first columns of $(BA)^T$ and DC respectively and $w_0 = [1/2 \;\; 1/2 \;\; 1/2 \;\; 1/2]^T$ is the eigenvector of $\tilde{R}_X$ that corresponds to the largest eigenvalue λ_0. This leads to the following relations that need to be satisfied for optimality. Equivalently

$$A^T \begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} \qquad D \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}.$$

Note that each of the relations give four equations in sixteen unknows so we have large freedom in choosing A and D. You can verify that the following choice for A and D satisfy the requirement.

$$A = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/8 & 1/8 & 1/8 & 1/8 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad D = \begin{bmatrix} 1/6 & 0 & 1/6 & 1/6 \\ 1/6 & 0 & 1/6 & 1/6 \\ 1/6 & 0 & 1/6 & 1/6 \\ 1/6 & 0 & 1/6 & 1/6 \end{bmatrix}.$$

For this choice of A and D, we have

$$BA = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/8 & 1/8 & 1/8 & 1/8 \\ 0 & 0 & 0 & 0 \\ 3/8 & 3/8 & 3/8 & 3/8 \end{bmatrix} \qquad DC = \begin{bmatrix} 1/2 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 & 1/2 \end{bmatrix}.$$

(d) The transformation by BA projects $\bar{X}[n]$ on the space (line) spanned by w_0, the result of this projection is passed to the next stage and the transformation by CD constructs the best estimate of $\bar{X}[n]$ again in the space spanned by w_0.

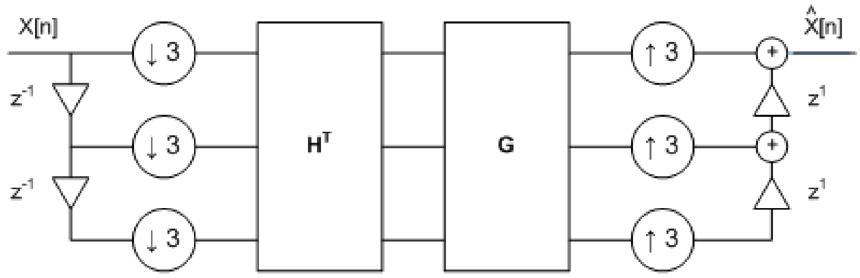
(e) We can obtain an equivalent system by choosing

$$H_0(z) = \frac{1}{2} + \frac{1}{2}z^{-1} + \frac{1}{2}z^{-2} + \frac{1}{2}z^{-3}$$

and

$$G_0(z) = \frac{1}{2} + \frac{1}{2}z + \frac{1}{2}z^{2} + \frac{1}{2}z^{3}.$$

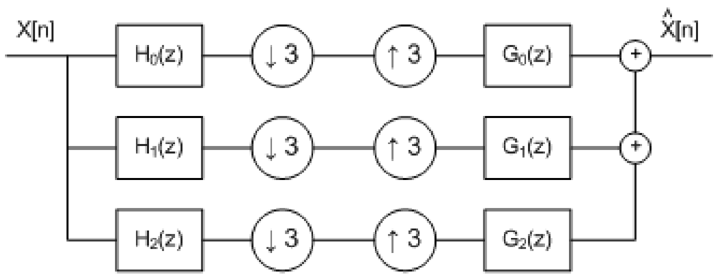
Exercise 58.
Consider:



with

$$\mathbf{H} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{bmatrix}$$

and

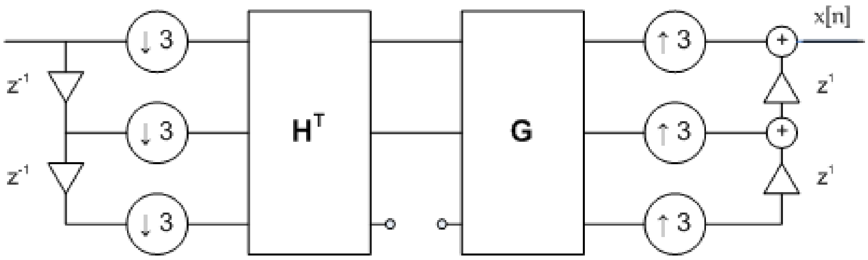


(a) Determine **G** in the first diagram in order to have perfect reconstruction.

(b) Determine $H_0(z)$, $H_1(z)$, and $H_2(z)$ such that the second diagram is equivalent to the first one.

Solution 58.

Exercise 59.
The process $x[n]$ is coded by using the following system:



The process $x[n]$ is stationary and has correlation $R_x[n] = \frac{1}{2^{|n|}}$. Suppose that **H** is given by:

$$\mathbf{H} \quad = \quad \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \mathbf{H}$$

Check if the given matrix **H** is one for which the average reconstruction error is minimum. (*Hint: there is a simple way to do this!*)

Exercise 60.
We have that

$$\mathbf{R_x} \quad = \quad \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & 1 \end{bmatrix}$$

If **H** is optimal, the first two columns should span the eigenspace corresponding to the two largest eigenvalues, while the last one should span the eigenspace corresponding to the minimum eigenvalue. This implies that the last column is an eigenvector.
In this case, we verify that

$$\mathbf{R_x} \cdot \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \quad = \quad \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \neq \lambda \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

This means that the last column is not an eigenvector, and therefore the transformation can not be optimal.

Exercise 61. Transform coding

Consider a zero-mean random variable $X[n]$ that we wish to encode in blocks of 4 samples, by keeping only the first 3 coefficients in the transform domain. For this purpose, we apply an orthogonal transform to $X[n]$ and encode the resulting coefficients $Y[n]$. The decoded coefficients $\hat{Y}[n]$ are then transformed back into reconstructed signal samples $\hat{X}[n]$.

(a) Represent this coding problem in a filterbank structure.

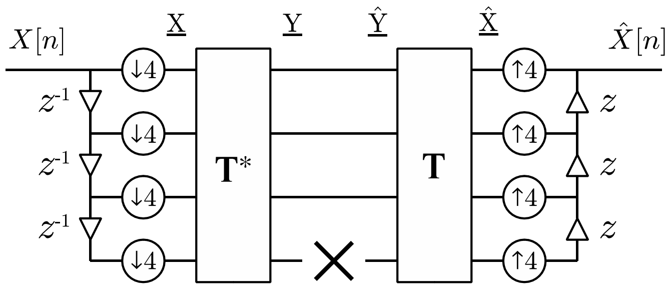
(b) Assuming the correlation of X is given by $R_X[l] = |l \mod 4 - 2|^2$, write down the correlation matrix $\mathbf{R}_X$. What type of matrix is $\mathbf{R}_X$? Is it diagonalizable?

(c) Determine the optimal transform matrix $\mathbf{T}$ that minimizes the reconstruction error. What is the reconstruction error? Is the process $\hat{X}$ stationary?

(d) Suppose the correlation is given instead by $R_X[l] = |l \mod 4 - 2|$. What is the reconstruction error? Is the process $\hat{X}$ stationary?

Solution 59. Transform coding

(a) The filterbank that performs the coding operation can be represented as follows.



The delay chain on the left generates a signal block of size 4, defined as

$$\underline{\mathrm{X}} = \begin{bmatrix} X[n] \\ X[n-1] \\ X[n-2] \\ X[n-3] \end{bmatrix},$$

which is then transformed into another domain by $\underline{\mathrm{Y}} = \mathbf{T}^*\underline{\mathrm{X}}$. The resulting transform coefficients are compressed (or coded) by discarding the last coefficient in $\underline{\mathrm{Y}}$, resulting in the approximation $\hat{\underline{\mathrm{Y}}}$. The inverse transform, given by $\hat{\underline{\mathrm{X}}} = \mathbf{T}\hat{\underline{\mathrm{Y}}}$, obtains an approximation of the input signal block, which is then recombined into a temporal signal $\hat{X}[n]$.

(b) By expanding $R_X[l]$ over l, we get:

$$\mathbf{R}_X = \begin{bmatrix} 4 & 1 & 0 & 1 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 1 & 0 & 1 & 4 \end{bmatrix}.$$

Notably, $\mathbf{R}_X$ is a circulant matrix, and therefore diagonalizable by a DFT matrix of size 4.

(c) The optimal transform matrix $\mathbf{T}$ is the KLT matrix, which contains the eigenvectors of $\mathbf{R}_X$ as columns sorted in descending order of the respective eigenvalues. Since $\mathbf{R}_X$ is diagonalized by the DFT matrix, $\mathbf{F}$, then the same contains the eigenvectors of $\mathbf{R}_X$ as columns. Accordingly,

$$\mathbf{F} = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix} = \begin{bmatrix} \mathbf{v}_0 & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}$$

and consequently

$$\mathbf{F}^*\mathbf{R}_X\mathbf{F} = \begin{bmatrix} 6 & & & \\ & 4 & & \\ & & 2 & \\ & & & 4 \end{bmatrix} = \begin{bmatrix} \lambda_0 & & & \\ & \lambda_1 & & \\ & & \lambda_2 & \\ & & & \lambda_3 \end{bmatrix}.$$

The only thing left is to sort the eigenvectors as stated, such that

$$\mathbf{T} = \begin{bmatrix} \mathbf{v}_0 & \mathbf{v}_1 & \mathbf{v}_3 & \mathbf{v}_2 \end{bmatrix}$$

and thus

$$\mathbf{T}^*\mathbf{R}_X\mathbf{T} = \begin{bmatrix} 6 & & & \\ & 4 & & \\ & & 4 & \\ & & & 2 \end{bmatrix}.$$

The reconstruction error is given by the sum of the individual errors associated to each discarded sample i, in which case

$$\sum_i \mathbf{v}_i^*\mathbf{R}_X\mathbf{v}_i = \mathbf{v}_2^*\mathbf{R}_X\mathbf{v}_2 = \lambda_2 = 4.$$

By discarding the last transform coefficient, we are introducing a non-linearity that makes the output process $\hat{X}$ non-stationary.

(d) The correlation matrix is given by

$$\mathbf{R}_X = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{bmatrix},$$

which is also a circulant matrix. In this case, we get

$$\mathbf{F}^*\mathbf{R}_X\mathbf{F} = \begin{bmatrix} 4 & & & \\ & 2 & & \\ & & 0 & \\ & & & 2 \end{bmatrix}.$$

The difference here is that the lowest eigenvalue is zero, and thus the reconstruction error is zero as well:

$$\mathbf{v}_2^*\mathbf{R}_X\mathbf{v}_2 = \lambda_2 = 0.$$

This means that the discarded transform coefficient is not necessary to reconstruct X, and thus $\hat{X}[n] = X[n]$ for all n. Moreover, the output process $\hat{X}$ is stationary.

Exercise 62. Z TRANSFORM
A causal and stable filter has the following Z transform.

$$H(z) = \frac{1 - 4\cos\frac{\pi}{4}z^{-1} + 4z^{-2}}{1 - \frac{1}{2}\cos\frac{3\pi}{4}z^{-1} + \frac{1}{16}z^{-2}}$$

(a) Find all poles and zeros and represent them on the z-plane.

Coding Using Matlab, plot the pole-zero diagram and verify your answer for the previous step.

(b) Determine a strictly minimum phase filter $H_m(z)$ and an all-pass filter $H_a(z)$ such that $H(z) = H_m(z)H_a(z)$.

Coding Using Matlab, plot the pole-zero diagrams for $H_m(z)$ and $H_a(z)$

(c) Determine the inverse filter $1/H_m(z)$ and, given that $1/H_m(z)$ is causal, explain why it is also stable and strictly minimum phase.

(d) Determine a filter $G(z)$ such that $H_m(z)G(z)$ is zero-phase. How can $G(z)$ be expressed in terms of $H_m(z)$?

Solution 60.

(a) We can factorize $H(z)$ as follows.

$$\begin{aligned} H(z) &= \frac{1 - 4\cos\frac{\pi}{4}z^{-1} + 4z^{-2}}{1 - \frac{1}{2}\cos\frac{3\pi}{4}z^{-1} + \frac{1}{16}z^{-2}} \\ &= \frac{1 - 2e^{j\frac{\pi}{4}}z^{-1} - 2e^{-j\frac{\pi}{4}}z^{-1} + 4z^{-2}}{1 - \frac{1}{4}e^{j\frac{3\pi}{4}}z^{-1} - \frac{1}{4}e^{-j\frac{3\pi}{4}}z^{-1} + \frac{1}{16}z^{-2}} \\ &= \frac{\left(1 - 2e^{j\frac{\pi}{4}}z^{-1}\right)\left(1 - 2e^{-j\frac{\pi}{4}}z^{-1}\right)}{\left(1 - \frac{1}{4}e^{j\frac{3\pi}{4}}z^{-1}\right)\left(1 - \frac{1}{4}e^{-j\frac{3\pi}{4}}z^{-1}\right)} \end{aligned}$$

from which we easily conclude that the poles are at $z = \frac{1}{4}e^{\pm j\frac{3\pi}{4}}$ and the zeros at $z = 2e^{\pm j\frac{\pi}{4}}$.

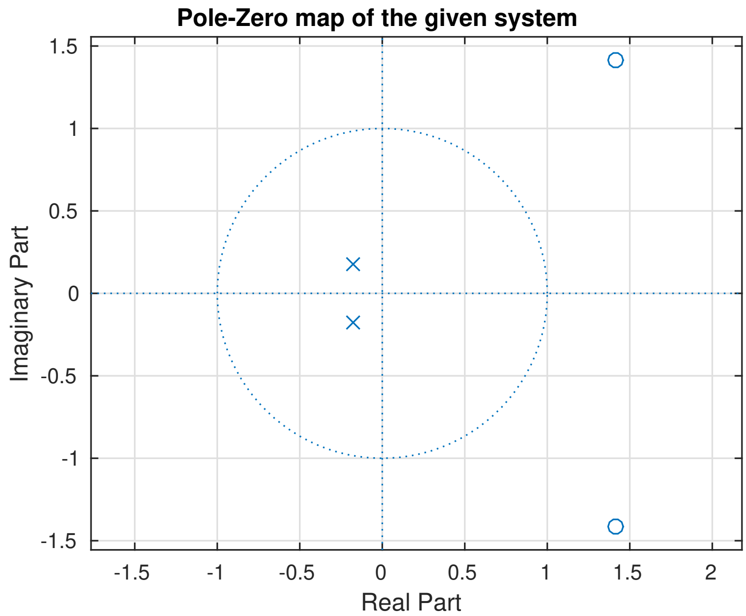
Matlab Using the *pzmap* and *zplane* commands in Matlab, the following pole-zero plot can easily be obtained:

(b) We can decompose $H(z)$ with the following procedure.

i Assign the poles and zeros inside the unit circle to $H_m(z)$ and the zeros on the outside to $H_a(z)$;

ii Make $H_a(z)$ an all-pass filter by adding poles that are conjugate reciprocals of the zeros;

iii Cancel the effect of the additional poles by adding zeros to $H_m(z)$ with the same values.



The result is as follows.

$$H_a(z) = \frac{\left(1 - 2e^{j\frac{\pi}{4}}z^{-1}\right)\left(1 - 2e^{-j\frac{\pi}{4}}z^{-1}\right)}{\left(1 - \frac{1}{2}e^{j\frac{\pi}{4}}z^{-1}\right)\left(1 - \frac{1}{2}e^{-j\frac{\pi}{4}}z^{-1}\right)}$$

$$H_m(z) = \frac{\left(1 - \frac{1}{2}e^{j\frac{\pi}{4}}z^{-1}\right)\left(1 - \frac{1}{2}e^{-j\frac{\pi}{4}}z^{-1}\right)}{\left(1 - \frac{1}{4}e^{j\frac{3\pi}{4}}z^{-1}\right)\left(1 - \frac{1}{4}e^{-j\frac{3\pi}{4}}z^{-1}\right)}$$

(c) The inverse filter is given by

$$\frac{1}{H_m(z)} = \frac{\left(1 - \frac{1}{4}e^{j\frac{3\pi}{4}}z^{-1}\right)\left(1 - \frac{1}{4}e^{-j\frac{3\pi}{4}}z^{-1}\right)}{\left(1 - \frac{1}{2}e^{j\frac{\pi}{4}}z^{-1}\right)\left(1 - \frac{1}{2}e^{-j\frac{\pi}{4}}z^{-1}\right)}$$

Since the resulting filter is causal and the outer pole is $\frac{1}{2}e^{\pm j\frac{\pi}{4}}$, then the region of convergence $|z| > \frac{1}{2}$ includes the unit circle, which implies that the filter is stable. Moreover, all poles and zeros are inside the unit circle, and thus the filter is strictly minimum phase.

(d) For $H_m(z)G(z)$ to be zero-phase, all poles and zeros must exist in conjugate reciprocal pairs. Thus, since

$$H_m(z) = \frac{\left(1 - \frac{1}{2}e^{j\frac{\pi}{4}}z^{-1}\right)\left(1 - \frac{1}{2}e^{-j\frac{\pi}{4}}z^{-1}\right)}{\left(1 - \frac{1}{4}e^{j\frac{3\pi}{4}}z^{-1}\right)\left(1 - \frac{1}{4}e^{-j\frac{3\pi}{4}}z^{-1}\right)}$$

then

$$G(z) = \frac{\left(1-2e^{j\frac{\pi}{4}}z^{-1}\right)\left(1-2e^{-j\frac{\pi}{4}}z^{-1}\right)}{\left(1-4e^{j\frac{3\pi}{4}}z^{-1}\right)\left(1-4e^{-j\frac{3\pi}{4}}z^{-1}\right)}$$

Furthermore, we notice that

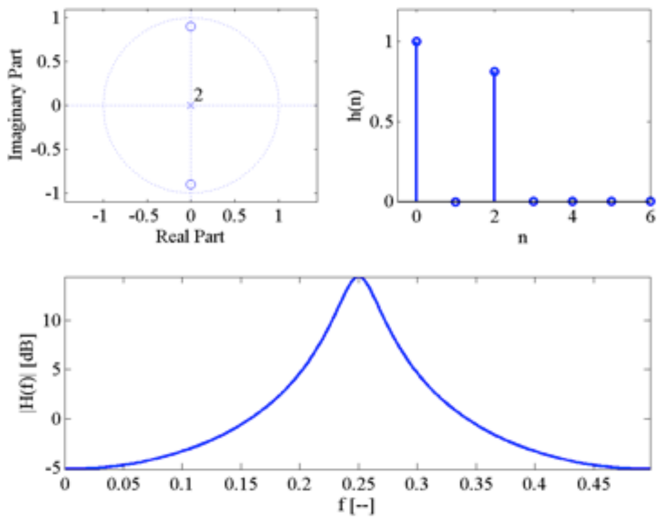
$$\begin{aligned}
H_m^*\left(\frac{1}{z^*}\right) &= \frac{\left(1-\frac{1}{2}e^{j\frac{\pi}{4}}z^*\right)^*\left(1-\frac{1}{2}e^{-j\frac{\pi}{4}}z^*\right)^*}{\left(1-\frac{1}{4}e^{j\frac{3\pi}{4}}z^*\right)^*\left(1-\frac{1}{4}e^{-j\frac{3\pi}{4}}z^*\right)^*} \\
&= \frac{\left(1-\frac{1}{2}e^{-j\frac{\pi}{4}}z\right)\left(1-\frac{1}{2}e^{j\frac{\pi}{4}}z\right)}{\left(1-\frac{1}{4}e^{-j\frac{3\pi}{4}}z\right)\left(1-\frac{1}{4}e^{j\frac{3\pi}{4}}z\right)} \\
&= \frac{-\frac{1}{2}e^{-j\frac{\pi}{4}}z\cdot -\frac{1}{2}e^{j\frac{\pi}{4}}z\left(\frac{1}{-\frac{1}{2}e^{-j\frac{\pi}{4}}z}+1\right)\left(\frac{1}{-\frac{1}{2}e^{j\frac{\pi}{4}}z}+1\right)}{-\frac{1}{4}e^{-j\frac{3\pi}{4}}z\cdot -\frac{1}{4}e^{j\frac{3\pi}{4}}z\left(\frac{1}{-\frac{1}{4}e^{-j\frac{3\pi}{4}}z}+1\right)\left(\frac{1}{-\frac{1}{4}e^{j\frac{3\pi}{4}}z}+1\right)} \\
&= \frac{4\left(1-2e^{j\frac{\pi}{4}}z^{-1}\right)\left(1-2e^{-j\frac{\pi}{4}}z^{-1}\right)}{\left(1-4e^{j\frac{3\pi}{4}}z^{-1}\right)\left(1-4e^{-j\frac{3\pi}{4}}z^{-1}\right)} \\
&= 4G(z)
\end{aligned}$$

We can verify the condition for zero-phase as follows.

$$\begin{aligned}
H_m(z)G(z) &= H_m^*\left(\frac{1}{z^*}\right)G^*\left(\frac{1}{z^*}\right) \\
&= 4G(z)\cdot\frac{1}{4}H_m(z)
\end{aligned}$$

Exercise 63. One system or more than one system?
Considering the following 3 plots:

- A z-transform in the z-plane (upper left corner), where `o` denotes the zeros and `x` the poles (the poles can in this framework be neglected);
- An impulse response $h(n)$ in the time domain (upper right corner);
- Magnitude of the frequency response $|H(e^{j2\pi f})|$ in normalized frequencies (bottom).



According to the plots:

(a) Do the z-transform and the impulse response $h(n)$ correspond to the same system (that is, is the plot of the z-plane the plot of the z-transform of $h(n)$)? You can plot the z-transform in Matlab and compare the plots.

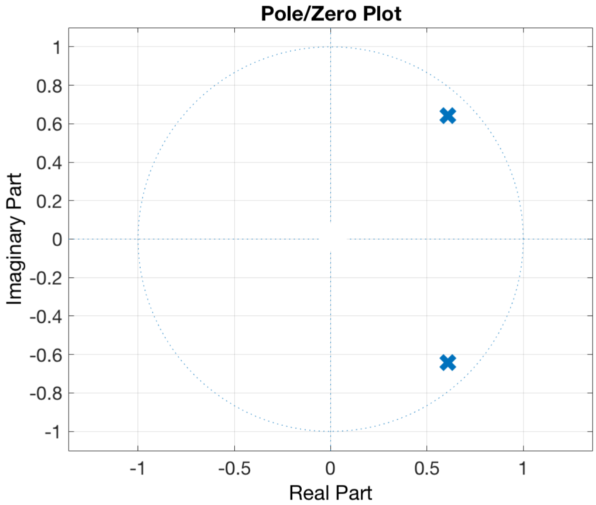
(b) Do the z-transform and the magnitude of the frequency response $|H(e^{j2\pi f})|$ correspond to the same system (that is, is $|H(e^{j2\pi f})|$ the absolute value on the unit circle of the z-transform represented in the z-plane plot)?

Solution 61. One system or more than one system?

(a) From the z-plane plot we see that the two zeros are $z_1 = j\alpha$ and $z_2 = -j\alpha$, where $0 < \alpha < 1$ (real). Consequently the z-transform has the form $\tilde{H}(z) = (1 - z_1 z^{-1})(1 - z_2 z^{-1}) = (1 - j\alpha z^{-1})(1 + j\alpha z^{-1}) = 1 + j\alpha z^{-1} - j\alpha z^{-1} + \alpha^2 z^{-2} = 1 + \alpha^2 z^{-2}$. The corresponding impulse response is therefore $\tilde{h}(0) = 1$, $\tilde{h}(1) = 0$, $\tilde{h}(2) = \alpha^2$, and $\tilde{h}(k) = 0$ for $k \geq 3$. Hence, $\tilde{h}(n) = h(n)$ and $\tilde{H}(z) = H(z)$.

(b) The z-transforms shows two zeros near the unit circle at normalized frequencies $f_1 = 0.25$ and $f_2 = -0.25$. Consequently the magnitude of the corresponding frequency response $|\tilde{H}(e^{j2\pi f})|$ should show a minimum of the frequency response at $f_1 = 0.25$ and a minimum at $f_2 = -0.25$. The plot of $|H(e^{j2\pi f})|$ (bottom) shows a maximum at $f_1 = 0.25$ (and for the symmetry of the spectrum, a maximum at $f_2 = -0.25$). Consequently, $|H(e^{j2\pi f})| \neq |H(e^{j2\pi f})|$ and the plot of the z-transform and the plot of the magnitude of the frequency response $|H(e^{j2\pi f})|$ do not correspond to the same system.

Exercise 64. A SIMPLE SYSTEM (4 PTS)
The figure below depicts the poles of a causal LTI system $P(z)$. the two poles have magnitude $a = 0.9$ and phase $\varphi = \pm\pi/4$.



(a) Is the system $P(z)$ stable?

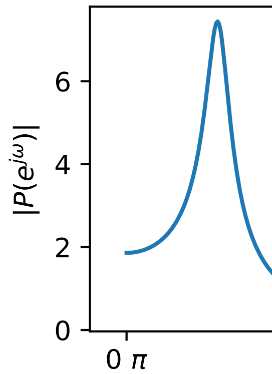
(b) Sketch the magnitude of the transfer function $\left|P(e^{jw})\right|$.

(c) Is the inverse system $H(z) = 1/P(z)$ stable?

(d) Give the impulse response $h(n)$ of $H(z)$.

Solution 62. A SIMPLE SYSTEM

(a) The system $P(z)$ is unstable, because it has poles inside unit circle.



(b) The magnitude $\left|P(e^{jw})\right|$ will have to peaks around $\pi/4$ and $7\pi/4$, as shown below:

(c) The inverse system $H(z) = 1/P(z)$ is stable because it has only zeros, and no poles inside the unit circle.

(d) We can write the frequency response of the system as:

$$P(z) = \frac{1}{(1 - p_0 z^{-1})(1 - \overline{p_0} z^{-1})} = \frac{A}{(1 - p_0 z^{-1})} + \frac{B}{(1 - \overline{p_0} z^{-1})}, \tag{6.5}$$

where $p_0 = 0.9e^{j\pi/4}$, $\overline{p_0}$ is it's conjugate and A and B are unknown constants we would like to recover. Each of the fractions can be expanded into a sum as follows:

$$\frac{1}{(1 - p_0 z^{-1})} = \sum_{n=0}^{\infty} p_0^n z^{-n},$$

and we can read that the impulse response has to be $h_{\text{part}}[n] = p_0^n$ (for this fraction only). Therefore, the total impulse response is:

$$h[n] = Ap_0^n + B\overline{p_0}^n.$$

By comparing the fractions in (6.5), we get that $A = \frac{p_0}{p_0 - \overline{p_0}} = e^{-j\pi/4}/\sqrt{2}$ and $B = \frac{\overline{p_0}}{\overline{p_0} - p_0} = e^{j\pi/4}/\sqrt{2}$, so we get

$$h[n] = \frac{0.9^n}{\sqrt{2}}(e^{-j\pi/4}e^{jn\pi/4} + e^{j\pi/4}e^{-jn\pi/4}) = 0.9^n\sqrt{2}\cos(\pi(n-1)/4)$$

Chapter 7

Markov Chains and Maximum Likelihood

Exercise 65. Hidden Markov Models and Likelihood
In this exercise, we will numerically model 2D Gaussian HMMs and generate sequences from them. Then, we will use the Forward algorithm to compute the likelihood that some sequence was generated by a given HMM. You can use the Matlab `struct` file named `data.mat`, which contains a couple HMMs stored as `struct`, each containing the mean values, variance values, and transition probabilities values.
Some Matlab helper functions are provided on the Moodle:

- `(y, state_sequence) = generate_hmm(hmm)`: generate a sequence of values `y` as well as a states sequence, from one HMM `struct` given as argument.
- `plot_sequence(y, state_sequence)`: plot the sequence `y` against time.
- `plot_sequence_2d(y, state_sequence)`: plot the sequence in 2 dimensions.
- `(likelihood) = forward_recursion(y, hmm)`: compute the likelihood that the sequence `y` was generated by HMM `hmm`.

(a) The HMM we are going to generate will have two additional states, which you could *start* and *stop*. Look at the provided data and identify those states. How can you interpret the probability of transition *from* the start state? Why the transition probability from the stop state

(b) Since the model for the HMM are 2D Gaussians, what are the dimensions of the *means matrix* and the *variance matrix*? How many of these matrices do we need?

(c) We want to generated HMMs with 3 states. How many transition probabilities do we need? What are the conditions on the transition matrix for it to be valid?

(d) Create a HMM model with 3 states with the programming language of your choice, this means, generate one mean matrix and one variance matrix for each state, as well as a transition matrix for the HMM. Alternatively, you can choose to load the `data.mat` which contains several ready-to-use HMMs. The means and variance matrix together with the transition probabilities are called the parameters of the HMM, and are denoted with Θ.

(e) Generate a sequence with the HMM. You can use the function `(y, state_sequence) = generate_hmm(hmm)` to help you. Why does the function return 2 sequences? What is the difference between these sequences? If only the sequence `state_sequence` was generated, would we still call this model a HMM?

(f) Try to plot the output sequence `y` against time (that is, the x axis is time, and the y axis is either the first dimension of `y`, either the second, so we need 2 plots to represent the sequence.) You can use `plot_sequence(y, state_sequence)` to help you do that. What do you observe? Try to change the means, variances, and transition probabilities and observe what changes on the plot.

(g) Now try to plot the sequence on a single plot, without caring of time progression, but instead using one axis per dimension of the sequence `y`. The function `plot_sequence_2d(y, state_sequence)` can help you do that, and also plots ellipses around the clusters of points. What do these ellipses correspond to? Again, try to change the means, variances, and transition probabilities and observe what changes on the plot.

(h) Create a second HMM model.

We now want to compute how likely some sequence `y` was generated from one HMM or another. We are going to do this using **Forward algorithm**, which is very well described on Wikipedia).

(a) Use the Forward algorithm to compute the likelihood of 2 sequences with the 2 HMM you generated (you can use sequences you've generated with the HMM, and check that each sequence as a higher likelihood to be generated of the HMM that indeed generated it). You can use `forward_recursion(y, hmm)` to compute the (log of the) likelihood. You can also test the likelihood of some other sequences given in the `data.mat` file.

(b) If you used the sequences you generated yourself, which sequence is the most likely to be generated from which HMM? Why?

Solution 63. HIDDEN MARKOV MODELS AND LIKELIHOOD
Solution is available in Matlab (in .zip).

Exercise 66.
In Markov chains, the probability that the chain is in a particular state k is given by π_k and the distribution of probabilities for states $1 \leq k \leq n$ can be represented by a vector π such that $0 \leq \pi_k \leq 1$, $\forall k$ and $\sum_{k=1}^{n} \pi_k = 1$. Markov chain is characterized by a matrix $\mathbf{M}$ such that the probability distribution π_{m+1} at instant $m+1$ can be computed from the distribution π_m at instant m as follows:

$$\pi_{m+1} = \mathbf{M}\pi_m \, .$$

The matrix $\mathbf{M} = [m_{ij}]$ has the following properties: $0 \leq m_{ij} \leq 1$ and $\sum_{i=1}^{n} m_{ij} = 1$, $\forall j$. An equilibrium distribution of the Markov chain π^e is a solution of the eigenvalue equation:

$$\mathbf{M}\pi^e = \pi^e$$

(a) Let $\|.\|_1$ be the 1-norm, defined as follows:

$$\|\mathbf{x}\|_1 = |x_1| + \ldots + |x_n|, \qquad \mathbf{x} \in \mathbb{C}^n \, .$$

Show that $\|\mathbf{M}\pi\|_1 = \|\pi\|_1$ for any probability distribution π.

(b) Does the sequence π_m always (for any initial vector π_0 with the above properties) converge to an equilibrium distribution π^e?

Solution 64.

(a)

$$\begin{aligned} \|\mathbf{M}\pi\|_1 &= \sum_{i=1}^{n} |(\mathbf{M}\pi)_i| = \sum_{i=1}^{n} |(\sum_{j=1}^{n} \pi_j \mathbf{m}_j)_i| = \sum_{i=1}^{n} |\sum_{j=1}^{n} \pi_j m_{ij}| \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_j m_{ij} = \sum_{j=1}^{n} \pi_j \sum_{i=1}^{n} m_{ij} = \sum_{j=1}^{n} \pi_j \cdot 1 = \sum_{j=1}^{n} |\pi_j| \\ &= \|\pi\|_1 \, . \end{aligned}$$

(b) No, it does not. A counterexample are the following matrix and vector

$$\mathbf{M} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \pi_{\mathbf{0}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{7.1}$$

It is easy to verify that the sequence does not converge.

Exercise 67. Markov Process
Let $\{X(t),\, t \geq 0\}$ be a stochastic process with state space $S = \{0, 1, \ldots, \infty\}$. Suppose that the two following conditions hold:

(a) $P(X(0) = 0) = 1$

(b) $\{X(t),\, t \geq 0\}$ has independent increments. That is, for every n and for every $0 \leq t_1 < \ldots < t_n$, $\{Y_i = X(t_i) - X(t_{i-1}),\, i = 1, \ldots, n\}$ are independently distributed from one another.

Show that $X(t)$ is a Markov process.

Solution 65. Markov Process
In order to prove that the process $X(t)$ is a Markov process, it suffices to prove that for every n, for every $0 \leq t_1 < \ldots < t_n$ and for every $x_1, \ldots, x_n \in S$, the following conditional probability:

$$P\left(X(t_n) = x_n \middle|\, X(t_{n-1}) = x_{n-1}, \ldots, X(t_1) = x_1\right) \tag{7.2}$$

should be equal to

$$P\left(X(t_n) = x_n \middle|\, X(t_{n-1}) = x_{n-1}\right) . \tag{7.3}$$

The probability given in (7.3) can be transformed in the following way:

$$\begin{aligned}
& P(X(t_n) = x_n |\, X(t_{n-1}) = x_{n-1}) \\
= \;& P(X(t_{n-1}) + Y_n = x_n |\, X(t_{n-1}) = x_{n-1}) \\
= \;& P(x_{n-1} + Y_n = x_n |\, X(t_{n-1}) = x_{n-1}) \\
= \;& P(Y_n = x_n - x_{n-1} |\, X(t_{n-1}) = x_{n-1})\,.
\end{aligned} \tag{7.4}$$

On the other hand, using the relation between the random variables $X(t_n), \ldots, X(t_1)$ and $Y_n, \ldots, Y_1$, the probability given in (7.2) can be transformed in the following way:

$$\begin{aligned}
& P(X(t_n) = x_n |\, X(t_{n-1}) = x_{n-1}, \ldots, X(t_1) = x_1) \\
= \;& P(X(t_{n-1}) + Y_n = x_n |\, X(t_{n-1}) = x_{n-1}, \ldots, X(t_1) = x_1) \\
= \;& P(Y_n = x_n - x_{n-1} |\, X(t_{n-1}) = x_{n-1}, \ldots, X(t_1) = x_1) \\
= \;& P(Y_n = x_n - x_{n-1} |\, X(t_{n-1}) = x_{n-1}, Y_{n-1} = x_{n-1} - x_{n-2}, \ldots, Y_1 = x_1) \\
= \;& P(Y_n = x_n - x_{n-1} |\, X(t_{n-1}) = x_{n-1})\,.
\end{aligned} \tag{7.5}$$

Since the final expressions in (7.4) and (7.5) are equal, the process $X(t)$ is a Markov process.

Exercise 68. Markov chains

(a) How many parameters are there in a *discrete* Markov chain with 3 states, all connected with each other (and to themselves)? What is the type (the name) of the parameters?

(b) How many parameters are there in a *hidden* 1-D Gaussian Markov model with 3 states, all connected with each other (and to themselves)? What is the type (the name) of the parameters?

Solution 66.

Exercise 69. Markov Chain
In this exercise you will generate Markov chain with 3 states. This could be for example the state of the weather: sunny, snowy and rainy. We can write transition probabilities in the matrix form:

$$\begin{bmatrix} 0.8 & 0.2 & 0.1 \\ 0.1 & 0.6 & 0.2 \\ 0.1 & 0.2 & 0.7 \end{bmatrix} \tag{7.6}$$

where the probabilities of the states at step $n+1$ given at step n the process was in the state i is defined by i-th column.

(a) *Warm up exercise:* Draw the diagram depicting this Markov Chain and it's transition probabilities.

(b) Generate a realisation of this Markov Chain in Python or Matlab, picking any state you like as the initial state. You can use provided Jupyter Notebook.

(c) Is your process (with a fixed starting point) stationary? What is the probability distribution after one, two and five steps? You can calculate the distribution in Python/Matlab.

(d) What equations the initial probabilities have to satisfy for the process to be stationary?

(e) Calculate such inital probabilities in Python/Matlab. Do it twice using different methods: using eigenvalue decomposition and estimate using your realisation of the process. Give example situations where you would use each method.

(f) Assume now that you don't know the the transition probabilities of your Markov Chain. Formulate the Maximum Likelihood as optimisation problem. What method would you use to solve it?

(g) *Additional Question*: Using the generated realisations of your signal solve your problem using `solve` in Matlab or `scipy.optimize.minimize` in Python. Is your estimation accurate?

Solution 67. Markov chain
Solution is available in Jupyter Notebook

Exercise 70. Markov Chain and Friends
Hidden Markov Model is Markov Chain observed through some additional random variable, for example through added noise.

(a) Explain that Gaussian Mixture model is a special case of Hidden Markov Model. What assumption do you have to add about Markov Chain?

(b) Explain that $Y[n]$ from the first part of the exercise previous exercise is a special case of Hidden Markov Model. What assumption do you have to add about Markov Chain?

(c) What about $Y[n]$ with a prior on $X[n]$?

(d) Pick and solve one of the Final Exam questions: **2018 2.C** or **2015 3.B**

Solution 68. Markov Chain and Friends

(a) If you take a (silly) special case where the states in the Markov Chain are i.i.d., that is when the transition probabilities don't depend on the previous state, and the Hidden Markov Model obtained by adding Gaussian noise to the states also a Gaussian Mixture Model

(b) We can think about the model from the previous exercise as single class Gaussian Mixture Model or in other words as a HMM with constant Markov Chain.

(c) On the other hand, if we have a *continuous* prior on A, this model becomes different form HMM, because we have a continuous and not discrete hidden state.

Exercise 71.

Suppose the signal $X[n]$ is constant, with amplitude A. Furthermore, let the observed signal $Y[n]$ contain a noisy version of the signal $X[n]$

$$Y[n] = X[n] + W[n]\,,$$

where $W[n]$ is a zero-mean white Gaussian noise with variance σ^2.

Assuming N observations of $Y[n]$ are given ($Y[0], \ldots, Y[N-1]$), give the *maximum likelihood estimation* of the parameter

$$\theta = \begin{pmatrix} A \\ \sigma^2 \end{pmatrix}$$

containing the unknown amplitude A and noise variance σ^2.

Solution 69.

First, it can be observed that since $X[n]$ is a constant ($X[n] = A$) and $W[n]$ is iid with distribution $\mathcal{N}(0, \sigma^2)$, the process $Y[n]$ is also iid, with distribution $\mathcal{N}(A, \sigma^2)$. In other words, knowing the value of $\theta = \begin{pmatrix} A \\ \sigma^2 \end{pmatrix}$, the joint distribution of N samples $Y[0], \ldots, Y[N-1]$ is given by

$$\begin{aligned} p(Y[1], \ldots, Y[n]|\theta) &\overset{iid}{=} \prod_{i=0}^{N-1} p(Y[i]|\theta) \\ &= \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y[i]-A)^2}{2\sigma^2}}\,. \end{aligned}$$

Furthermore, according to the definition, the likelihood function associated to the probability density function of the observed data is given by

$$\begin{aligned} \mathcal{L}(\theta) &\triangleq p(Y[1], \ldots, Y[n]|\theta) \\ &= \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y[i]-A)^2}{2\sigma^2}}\,, \end{aligned}$$

and its logarithm, the *log-likelihood*, is

$$\begin{aligned} \mathcal{L}^*(\theta) &= \ln \mathcal{L}(\theta) \\ &= -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \sum_{i=0}^{N-1} \frac{(Y[i]-A)^2}{2\sigma^2}\,. \end{aligned}$$

Maximizing the likelihood function is equivalent to maximizing the log-likelihood, which can be done by setting to zero the partial derivatives of $\mathcal{L}^*(\theta)$ with respect to A and σ^2. It gives:

$$\begin{aligned}
\frac{\partial \mathcal{L}^*(\theta)}{\partial A} = 0 \quad &\Leftrightarrow \quad 2\sum_{i=0}^{N-1} \frac{Y[i] - A}{2\sigma^2} = 0 \\
&\Rightarrow \quad \hat{A} = \frac{1}{N}\sum_{i=0}^{N-1} Y[i]\,; \\
\frac{\partial \mathcal{L}^*(\theta)}{\partial \sigma^2} = 0 \quad &\Rightarrow \quad -\frac{N}{2\sigma^2} + \sum_{i=0}^{N-1} \frac{(Y[i] - A)^2}{2\sigma^4} = 0 \\
&\Rightarrow \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=0}^{N-1} (Y[i] - \hat{A})^2\,.
\end{aligned}$$

Exercise 72. Maximum Likelihood

Suppose the signal $X[n]$ is constant, with amplitude A. Furthermore, let the observed signal $Y[n]$ contain a noisy version of the signal $X[n]$

$$Y[n] = X[n] + W[n]\,,$$

where $W[n]$ is a zero-mean white Gaussian noise with variance σ^2.

(a) Assuming N observations of $Y[n]$ are given ($Y[0], \ldots, Y[N-1]$), give the *maximum likelihood estimation* of the parameter

$$\theta = \begin{pmatrix} A \\ \sigma^2 \end{pmatrix}$$

containing the unknown amplitude A and noise variance σ^2.

(b) Download data file `MaximumLikelihood.csv` from Moodle. It was generated from Y. Estimate θ *using the formulas you derived in the previous part.*

(c) Is your estimator unbiased? Is the estimator in the programming language of your choice unbiased?

Solution 70. Maximum Likelihood

(a) First, it can be observed that since $X[n]$ is a constant ($X[n] = A$) and $W[n]$ is iid with distribution $\mathcal{N}(0, \sigma^2)$, the process $Y[n]$ is also iid, with distribution $\mathcal{N}(A, \sigma^2)$. In other words, knowing the value of $\theta = \begin{pmatrix} A \\ \sigma^2 \end{pmatrix}$, the joint distribution of N samples $Y[0], \ldots, Y[N-1]$ is given by

$$\begin{aligned} p(Y[1], \ldots, Y[n]|\theta) &\overset{iid}{=} \prod_{i=0}^{N-1} p(Y[i]|\theta) \\ &= \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y[i]-A)^2}{2\sigma^2}}\,. \end{aligned}$$

Furthermore, according to the definition, the likelihood function associated to the probability density function of the observed data is given by

$$\begin{aligned} \mathcal{L}(\theta) &\triangleq p(Y[1], \ldots, Y[n]|\theta) \\ &= \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y[i]-A)^2}{2\sigma^2}}\,, \end{aligned}$$

and its logarithm, the *log-likelihood*, is

$$\begin{aligned} \mathcal{L}^*(\theta) &= \ln \mathcal{L}(\theta) \\ &= -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \sum_{i=0}^{N-1} \frac{(Y[i]-A)^2}{2\sigma^2}\,. \end{aligned}$$

Maximizing the likelihood function is equivalent to maximizing the log-likelihood, which can be done by setting to zero the partial derivatives of $\mathcal{L}^*(\theta)$ with respect to A and σ^2. It gives:

$$\begin{aligned}
\frac{\partial \mathcal{L}^*(\theta)}{\partial A} = 0 \quad &\Leftrightarrow \quad 2\sum_{i=0}^{N-1} \frac{Y[i] - A}{2\sigma^2} = 0 \\
&\Rightarrow \quad \hat{A} = \frac{1}{N}\sum_{i=0}^{N-1} Y[i]\,; \\
\frac{\partial \mathcal{L}^*(\theta)}{\partial \sigma^2} = 0 \quad &\Rightarrow \quad -\frac{N}{2\sigma^2} + \sum_{i=0}^{N-1} \frac{(Y[i] - A)^2}{2\sigma^4} = 0 \\
&\Rightarrow \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=0}^{N-1}(Y[i] - \hat{A})^2\,.
\end{aligned}$$

(b) On this particular realisation of the process, mean of the samples is ~ 3.17. The variance using your estimator is ~ 2.26. If you use Python, then the build in `np.var` also gives you ~ 2.26, but in Matlab you get ~ 2.27 if you use `var`.

(c) The estimator of the mean is unbiased but the estimator of variance is biased. You can infer it by looking at the unbiased estimator introduced in the lecture:

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=0}^{N-1}(Y[i] - \hat{A})^2$$

and noticing that your estimator differs by a factor $\frac{N-1}{N}$.

Alternatively, you can calculate the expected value of $\hat{\sigma}$, but is a long calculation. We have to be careful because estimator of σ depends on the estimator the mean, which is not independent of the data, and thus we have to replace it with the mean of samples.

Before we do this, let us consider the implementations of variance estimators in Python and Matlab. In the first one, the default estimator is biased, exactly the one we have calculated, and you can observe that the values of the estimators match (see previous point). In Matlab however, the estimator is unbiased, and we can see that on our data it's value is slightly higher. In both those languages, you can pass a flag that will change the behaviour of the estimator.

Exact calculation of the expectation of the estimator:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{N}\sum_{i=0}^{N-1}(Y[i] - \hat{A})^2 \\
&= \frac{1}{N}\sum_{i=0}^{N-1}(Y[i] - \frac{1}{N}\sum_{k=0}^{N-1} Y[k])^2 \\
&= \frac{1}{N}\sum_{i=0}^{N-1}(\frac{N-1}{N}Y[i] - \frac{1}{N}\sum_{k\neq i}^{N-1} Y[k])^2
\end{aligned}$$

Now, the first component of the sum is independent of the second one. More over, the expression under the sum does not depend on the choice of i and we can write:

$$\begin{aligned}
\mathbb{E}(\hat{\sigma}^2) &= \left(\frac{N-1}{N}\right)^2 \mathbb{E}Y_0^2 - 2\frac{N-1}{N}\mathbb{E}Y_0\mathbb{E}\left(\sum_{k=1}^{N-1}\frac{1}{N}Y_k\right) + \mathbb{E}\left(\sum_{k=1}^{N-1}\frac{1}{N}Y_k\right)^2 \\
&= \left(\frac{N-1}{N}\right)^2 \left(\mathbb{E}Y_0^2 - 2(\mathbb{E}Y_0)^2\right) + \frac{1}{N^2}\mathbb{E}\left(\left(\sum_{k=1}^{N-1}Y_k\right)\left(\sum_{j=1}^{N-1}Y_j\right)\right) \\
&= \left(\frac{N-1}{N}\right)^2 \left(\mathbb{E}Y_0^2 - 2(\mathbb{E}Y_0)^2\right) + \frac{1}{N^2}\left(\left((N-1)^2 - N\right)(\mathbb{E}Y_0)^2 + N\mathbb{E}Y_0^2\right),
\end{aligned}$$

where the last equality comes form considering $j = k$ and $j \neq k$ separately. Just by rearranging the terms we get:

$$\begin{aligned}
\mathbb{E}(\hat{\sigma}^2) &= \frac{(N-1)^2}{N^2}\left(\mathbb{E}Y_0^2 - (\mathbb{E}Y_0)^2\right) + \frac{1}{N}\left(\mathbb{E}Y_0^2 - (\mathbb{E}Y_0)^2\right) \\
&= \frac{N-1}{N}\mathrm{Var}(Y) \neq \mathrm{Var}(Y)
\end{aligned}$$

So the estimator is biased.

Exercise 73. PRIOR, POSTERIOR, LIKELIHOOD

Suppose the signal $X[n]$ is constant, with amplitude A. Furthermore, let the observed signal $Y[n]$ contain a noisy version of the signal $X[n]$

$$Y[n] = X[n] + W[n]\,,$$

where $W[n]$ is a zero-mean white Gaussian noise with variance $\sigma^2 = 1$.

(a) Assuming N observations of $Y[n]$ are given $(Y[0], \ldots, Y[N-1])$, give the *maximum likelihood estimation* of amplitude A

(b) Download data file `Review_3.csv` from Moodle. It was generated from Y. Estimate θ *using the formulas you derived in the previous part.*

(c) Now assume that A is in itself a random variable, with prior distribution $A \sim \mathcal{N}(0, \sigma_A^2)$. What is MAP (maximum a posteriori) estimator of A?

(d) How can you relate *prior, posterior* and *likelihood* in one equation?

(e) Assuming different σ_A^2 (for example $0.001, 1$ and 100), estimate from the data A using MAP. What do you observe?

Solution 71.

(a) First, it can be observed that since $X[n]$ is a constant ($X[n] = A$) and $W[n]$ is iid with distribution $\mathcal{N}(0,1)$, the process $Y[n]$ is also iid, with distribution $\mathcal{N}(A,1)$. In other words, knowing the value of A, the joint distribution of N samples $Y[0], \ldots, Y[N-1]$ is given by

$$\begin{aligned} p(Y[1], \ldots, Y[n]|A) &\overset{iid}{=} \prod_{i=0}^{N-1} p(Y[i]|A) \\ &= \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{(Y[i]-A)^2}{2}}\,. \end{aligned}$$

Furthermore, according to the definition, the likelihood function associated to the probability density function of the observed data is given by

$$\begin{aligned} \mathcal{L}(A) &\triangleq p(Y[1], \ldots, Y[n]|A) \\ &= \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{(Y[i]-A)^2}{2}}\,, \end{aligned}$$

and its logarithm, the *log-likelihood*, is

$$\begin{aligned} \mathcal{L}^*(A) &= \ln \mathcal{L}(A) \\ &= -\frac{N}{2} \ln 2\pi - \sum_{i=0}^{N-1} \frac{(Y[i]-A)^2}{2}\,. \end{aligned}$$

Maximizing the likelihood function is equivalent to maximising the log-likelihood, which can be done by setting to zero the partial derivatives of $\mathcal{L}^*(A)$ with respect to A. It gives:

$$\frac{\partial \mathcal{L}^*(A)}{\partial A} = 0 \quad \Leftrightarrow \quad 2\sum_{i=0}^{N-1} \frac{Y[i]-A}{2} = 0$$
$$\Rightarrow \quad \hat{A} = \frac{1}{N}\sum_{i=0}^{N-1} Y[i]\,;$$

(b) For the data provided, the mean is 5.71.

(c) Our likelihood from the first part of the exercise have been calculated assuming that A is a parameter, so we have $p(Y[i]|A)$. We are now interested in posterior distribution, that is:

$$p(A|Y[i]).$$

We can calculate it using Bayes rule:

$$p(A|Y[i])p(Y[i]) = p(Y[i], A) = p(Y[i]|A)p(A)$$

In a more elaborate way, we can write it as:

$$\text{posterior} \times \text{evidence} = \text{likelihood} \times \text{prior}.$$

Since evidence is probability of data, it's constant for the given data and we can ignore it in the optimisation. We are then interested in maximising:

$$\begin{aligned} p(Y[1], \ldots, Y[n]|A)p(A) &= \left(\prod_{i=0}^{N-1} p(Y[i]|A)\right) p(A) \\ &= \left(\prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{(Y[i]-A)^2}{2}}\right) \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{A^2}{2\sigma_A^2}} \end{aligned}$$

similarly like with likelihood optimisation, it's easier to optimise the logarithm:

$$-\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln 2\pi\sigma_A^2 - \sum_{i=0}^{N-1} \frac{(Y[i]-A)^2}{2} - \frac{A^2}{2\sigma_A^2},$$

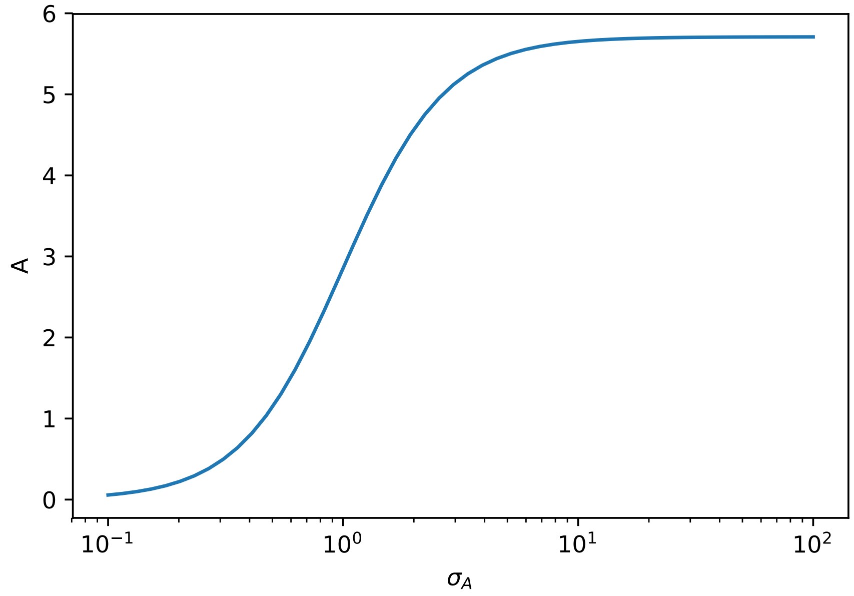
we want it's derivative to be equal to zero:

$$-\sum_{i=0}^{N-1} (Y[i]-A) + \frac{A}{\sigma_A^2} = 0 \tag{7.7}$$

And maximum a posterior estimator is:

$$\hat{A} = \frac{\sum_{i=0}^{N-1} Y[i]}{\left(N + \frac{1}{\sigma_A^2}\right)} \tag{7.8}$$

(d) Value of the estimator $\hat{A}$ depending on the variance of the prior σ_A^2. Small variance of the prior means we have a strong belief that A should be close to zero, and this belief is reflected in the value of the estimator, see the plot below.



6
5
4
3
2
1
0
A
10^{-1}
10^{0}
10^{1}
10^{2}
σ_A

Exercise 74. Poisson Process
Poisson process is the continues time process consisting with spikes only. There are two ways of generating it, and in this exercise we will see that they are equivalent.
The common definition of Poisson process is that the number of spikes k on the interval of length t has Poisson distribution:

$$p(k, t\lambda) = \frac{(t\lambda)^k e^{-t\lambda}}{k!} \tag{7.9}$$

On the interval $[0, t)$, given the number of spikes, the spikes positions $t_1, t_2, \ldots, t_k$ are distributed uniformly:

$$(t_1, t_2, \ldots, t_k) \sim U^k(0, t).$$

And similarly, on any interval $[t_0, t_0 + t)$:

$$(t_1, t_2, \ldots, t_k) \sim U^k(t_0, t_0 + t). \tag{7.10}$$

(a) *Warm up question*: Is the process $y[n]$ of the number of spikes in the interval $[n, n + 1)$ stationary?

(b) Calculate maximum likelihood estimator of parameter λ of Poisson Process. *Hint:* you can count the number of samples first.

(c) In Matlab or Python generate a realisation of the Poisson process re-estimate parameter λ. How long your interval should be for accurate estimation of λ? Does it matter if you average over multiple intervals? You can use the provided Jupyter Notebook.

Now we will generate the spike times t_i directly using exponential distribution on the *times of arrival*, $t_{i+1} - t_i$:

$$\begin{aligned} t_1 &\sim \exp(t, \tau) \\ t_{i+1} - t_i &\sim \exp(t, \tau) \end{aligned} \tag{7.11}$$

where the pdf of exponential distribution is

$$p(t, \tau) = \begin{cases} \tau e^{-\tau t} & t \geq 0, \\ 0 & t < 0. \end{cases}$$

(d) Generate realisations of this process for different τ.

(e) Use the maximum likelihood estimator from the previous part of the exercise to estimate Poisson process parameter λ.

(f) Plot estimated λ against τ. What is the dependence between those parameters? Is that what you would expect?

(g) *Additional question:* Prove that indeed the two discussed methods generate the same process (for the right choice of λ and τ).

Solution 72. Poisson process
Solution is available in Jupyter Notebook