Delft University of Technology
Faculty of Electrical Engineering, Mathematics, and Computer Science
Circuits and Systems Group

EE4C03 STATISTICAL DIGITAL SIGNAL PROCESSING AND MODELING

4 November 2020, 13:30–16:30

Block 1 (13:30-15:00)

Open book, strictly timed take-home exam. (Electronic) copies of the book and the course slides allowed. No other tools except a basic pocket calculator permitted.

Upload answers during 14:55–15:05

This block consists of three questions (25 points); more than usual, and this will be taken into account during grading. Answer in Dutch or English. Make clear in your answer how you reach the final result; the road to the answer is very important. Write your name and student number on each sheet.

Hint: Avoid losing too much time on detailed calculations, write down the general approach first.

Question 1 (8 points)

(a) Let $x(n) = 4 + v(n)$, where $v(n)$ is real-valued zero mean i.i.d. noise with variance $\sigma_v^2 = 1$. The signal is filtered by an FIR filter with impulse response

$$h(n) = \begin{cases} 1 & n = 0, \\ -\frac{1}{2} & n = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The output of the filter is $y(n) = h(n) * x(n)$.

Compute the mean value $m_y(n) = E\{y(n)\}$ and the autocovariance sequence $c_y(k)$.

(b) For a zero mean WSS process, let the autocorrelation sequence be given by

$$r(k) = A\cos(\omega_0 k)\,,$$

and construct the $N \times N$ correlation matrix $\mathbf{R}$.

Derive that the rank of $\mathbf{R}$ is equal to 2.

Hint: first write $\cos(\omega_0 k)$ *as the sum of two exponentials.*

(c) *True* or *False:* is this a valid autocorrelation matrix?

$$\mathbf{R} = \begin{bmatrix} 2 & 2 & 0 & 0 \\ 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

(Motivate your answer.)

(d) Suppose we have N samples of a zero mean WSS random process $x(n)$, but with one sample missing:

$$\mathbf{x} = [x(0), x(1), \cdots, x(n_0 - 1), x(n_0 + 1), \cdots, x(N-1)]^T$$

As usual, denote the correlation sequence by $r(k) = E\{x(n+k)x^*(n)\}$, and denote the correlation matrix $\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\}$.

Write down $\mathbf{R}$ in terms of $r(k)$. Is this a Toeplitz matrix? And can the usual correlation matrix corresponding to a sequence with no missing samples be reconstructed?

(e) *True* or *False:* if $y(n) = x_1(n) + x_2(n)$, and $x_1(n)$ is an AR(1) process and $x_2(n)$ is an AR(2) process, then $y(n)$ is an AR(3) process. (Motivate your answer.)

Solution

(a)

$$m_y(n) = m_x(n) \sum h_k = 4 \cdot \frac{1}{2} = 2\,.$$

$$\begin{aligned} c_x(k) &= \delta(k) \\ c_y(k) &= h(k) * h(-k) * c_x(k) \\ &= [\cdots \quad 0 \quad -\tfrac{1}{2} \quad \boxed{1\tfrac{1}{4}} \quad -\tfrac{1}{2} \quad 0 \quad \cdots] * [\cdots \quad 0 \quad \boxed{1} \quad 0 \quad \cdots] \\ &= [\cdots \quad 0 \quad -\tfrac{1}{2} \quad \boxed{1\tfrac{1}{4}} \quad -\tfrac{1}{2} \quad 0 \quad \cdots] \end{aligned}$$

Alternative computation: $P_y(z) = (1 - \frac{1}{2}z)(1 - \frac{1}{2}z^{-1}) \cdot 1 = 1\frac{1}{4} - \frac{1}{2}z - \frac{1}{2}z^{-1}$, with the same result.

(b) $r(k) = A\cos(\omega_0 k) = \frac{A}{2}e^{j\omega_0 k} + \frac{A}{2}e^{-j\omega_0 k}$.

The first term generates as covariance matrix

$$\begin{aligned} \mathbf{R}_1 &= \frac{A}{2}\begin{bmatrix} 1 & e^{-j\omega_0} & e^{-j2\omega_0} & \cdots & e^{-j(N-1)\omega_0} \\ e^{j\omega_0} & 1 & e^{-j\omega_0} & \cdots & e^{-j(N-2)\omega_0} \\ e^{j2\omega_0} & e^{j\omega_0} & 1 & \cdots & e^{-j(N-3)\omega_0} \\ \ddots & \ddots & \ddots & & \ddots \\ e^{j(N-1)\omega_0} & e^{j(N-2)\omega_0} & e^{j(N-3)\omega_0} & \cdots & 1 \end{bmatrix} \\ &= \frac{A}{2}\begin{bmatrix} 1 \\ e^{j\omega_0} \\ e^{j2\omega_0} \\ \vdots \\ e^{j(N-1)\omega_0} \end{bmatrix} [1 \;\; e^{-j\omega_0} \;\; e^{-j2\omega_0} \;\; \cdots \;\; e^{-j(N-1)\omega_0}] \end{aligned}$$

which is rank 1. Then, $\mathbf{R} = \mathbf{R}_1 + \mathbf{R}_1^H$ has rank 2.

(c) *False.* The matrix is Hermitian, Toeplitz, and $r(0) > |r(k)|$. But because $r(0) = 2$, $r(1) = 2$, the process is periodic and we should have $r(2) = 2$, $r(3) = 2, \cdots$. But $r(3) = 0$.

With Matlab, you could also compute the eigenvalues of $\mathbf{R}$ and note that they are not all semi-positive. You could also compute the determinant and note that it is negative.

(d)

$$\mathbf{R} = \left[\begin{array}{cccc|ccc} r(0) & r(1) & \cdots & r(n_0-1) & r(n_0+1) & \cdots & r(N-1) \\ r(1) & r(0) & \cdots & r(n_0-2) & r(n_0) & \cdots & r(N-2) \\ r(2) & r(1) & \cdots & r(n_0-3) & r(n_0-1) & \cdots & r(N-3) \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ r(n_0-1) & r(n_0-2) & \cdots & r(0) & r(2) & \cdots & r(N-n_0) \\ \hline r(n_0+1) & r(n_0) & \cdots & r(2) & r(0) & \cdots & r(N-n_0-2) \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ r(N-1) & r(N-2) & \cdots & r(N-n_0) & r(N-n_0-2) & \cdots & r(0) \end{array}\right]$$

It is not exactly a Toeplitz matrix as a row and column are missing. (E.g., the diagonal below the main diagonal has both $r(1)$ and $r(2)$.) The usual $\mathbf{R}$ can be recovered as $r(k)$ is known.

(e) *False,* in general. An AR(3) process is generated by a white noise process filtered with a third order AR filter. However, $x_1(n)$ and $x_2(n)$ are generated by independent white noise sequences. This is a different model.

Question 2 (8 points)

Let

$$x(n) = \sin(\omega_0 n + \phi_0)\,,$$

where ϕ_0 is a random variable, uniformly distributed in the range $[-\pi, \pi]$, and ω_0 is fixed.

(a) Show that $x(n)$ is zero mean, and derive that the autocovariance sequence of $x(n)$ is

$$r_x(k) = \tfrac{1}{2}\cos(\omega_0 k)\,.$$

(b) Prove that $x(n)$ satisfies the identity

$$x(n) - 2\cos(\omega_0)x(n-1) + x(n-2) = 0$$

Hint: $\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B)$.

(c) Explain that this implies that $x(n)$ is "*deterministic*" (predictable).

(d) For a specific ω_0, the first few samples of $r_x(k)$ are

$$r_x(0) = 0.5\,, \qquad r_x(1) = 0.25\,, \qquad r_x(2) = -0.25\,, \qquad r_x(3) = -0.5 \qquad \cdots$$

Using the Schur algorithm, compute, step by step, the reflection coefficients Γ_1 and Γ_2.

(e) Looking at $|\Gamma_2|$, what can you conclude?

(f) Suppose now that $x(n)$ is the sum of 2 sinusoids. What can you say about the reflection coefficients?

Solution

(a) This is shown in Hayes, p. 78

(b) Use $\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B)$ with $A = \omega_0(n-1)$, $B = \omega_0$:

$$\sin(\omega_0 n + \phi_0) = \cos(\omega_0)\sin(\omega_0(n-1) + \phi_0) + \sin(\omega_0)\cos(\omega_0(n-1) + \phi_0)$$

Also, with $A = \omega_0(n-1)$, $B = -\omega_0$:

$$\sin(\omega_0(n-2) + \phi_0) = \cos(\omega_0)\sin(\omega_0(n-1) + \phi_0)) - \sin(\omega_0)\cos(\omega_0(n-1) + \phi_0))$$

Summing both expressions gives the result. Alternatively,

$$\begin{aligned}
& x(n) - 2\cos(\omega_0)x(n-1) + x(n-2) \\
= \; & \frac{1}{2j}e^{j(\omega_0 n+\phi_0)} - \frac{1}{2j}e^{-j(\omega_0 n+\phi_0)} - (e^{j\omega_0} + e^{-j\omega_0})\left(\frac{1}{2j}e^{j(\omega_0(n-1)+\phi_0)} - \frac{1}{2j}e^{-j(\omega_0(n-1)+\phi_0)}\right) \\
& + \frac{1}{2j}e^{j(\omega_0(n-2)+\phi_0)} - \frac{1}{2j}e^{-j(\omega_0(n-2)+\phi_0)} \\
= \; & \frac{1}{2j}e^{j(\omega_0 n+\phi_0)} - \frac{1}{2j}e^{-j(\omega_0 n+\phi_0)} - \frac{1}{2j}e^{j\omega_0}e^{j(\omega_0(n-1)+\phi_0)} + \frac{1}{2j}e^{j\omega_0}e^{-j(\omega_0(n-1)+\phi_0)} \\
& - \frac{1}{2j}e^{-j\omega_0}e^{j(\omega_0(n-1)+\phi_0)} + \frac{1}{2j}e^{-j\omega_0}e^{-j(\omega_0(n-1)+\phi_0)} + \frac{1}{2j}e^{j(\omega_0(n-2)+\phi_0)} - \frac{1}{2j}e^{-j(\omega_0(n-2)+\phi_0)} \\
= \; & \frac{1}{2j}e^{j(\omega_0 n+\phi_0)} - \frac{1}{2j}e^{-j(\omega_0 n+\phi_0)} - \frac{1}{2j}e^{j(\omega_0 n+\phi_0)} + \frac{1}{2j}e^{-j(\omega_0(n-2)+\phi_0)} \\
& - \frac{1}{2j}e^{j(\omega_0(n-2)+\phi_0)} + \frac{1}{2j}e^{-j(\omega_0 n+\phi_0)} + \frac{1}{2j}e^{j(\omega_0(n-2)+\phi_0)} - \frac{1}{2j}e^{-j(\omega_0(n-2)+\phi_0)} \\
= \; & 0\,.
\end{aligned}$$

(c) Knowing two samples of $x(n)$ and ω_0, we can perfectly predict the entire sequence. If ω_0 is not known, we need 3 samples.

(d)

$$\begin{bmatrix} 0 & 0.5 \\ 0.25 & 0.25 \\ -0.25 & -0.25 \\ -0.5 & -0.5 \\ \vdots & \vdots \end{bmatrix} \underset{\text{shift}}{\Longrightarrow} \begin{bmatrix} 0 & 0 \\ 0.25 & 0.5 \\ -0.25 & 0.25 \\ -0.5 & -0.25 \\ \vdots & \vdots \end{bmatrix} \underset{\text{rotate, } \Gamma_1 = -\frac{0.25}{0.5}}{\Longrightarrow} \begin{bmatrix} 0 & 0 \\ 0 & 0.375 \\ -0.375 & 0.375 \\ -0.375 & 0 \\ \vdots & \vdots \end{bmatrix}$$

$$\underset{\text{shift}}{\Longrightarrow} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -0.375 & 0.375 \\ -0.375 & 0.375 \\ \vdots & \vdots \end{bmatrix} \underset{\text{rotate, } \Gamma_2 = \frac{0.375}{0.375}}{\Longrightarrow} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \end{bmatrix}$$

Thus, $\Gamma_1 = -\frac{1}{2}$ and $\Gamma_2 = 1$. At this point, the prediction error is zero and the recursion stops.

(e) $|\Gamma_2| = 1$, indicating that the process is perfectly predictable (no residual prediction error). It is marginally stable, resulting in an oscillation (no input is needed to sustain a sinusoid at the output).

Two reflection coefficients $\Rightarrow$ filter model has order 2.

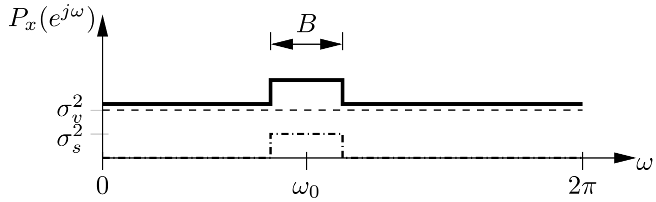
(f) In this case, the filter model has order 4, so $|\Gamma_4| = 1$.

Question 3 (9 points)

We are given 2000 samples of a single signal in noise,

$$x(n) = s(n) + v(n)$$

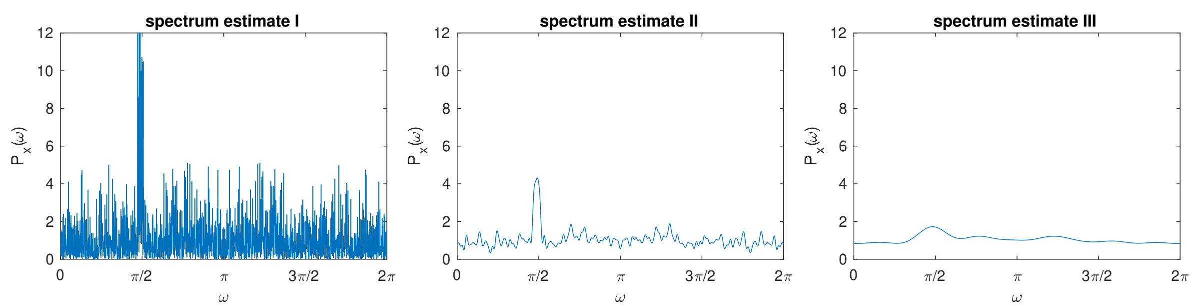
where $v(n)$ is zero mean white Gaussian noise with variance $\sigma_v^2 = 1$, while $s(n)$ is a weak astronomical signal, modeled as zero mean Gaussian noise with power spectrum $P_s(e^{j\omega})$, the dash-dot line in the figure. Based on physics, we expect the signal at $\omega_0 = 1.5$ rad and with a bandwidth $B = 0.16$ rad. We do not know its power in the passband.



We wish to detect the presence of the astronomical signal from a Bartlett spectrum estimate $\hat{P}_B(e^{j\omega})$. The $N = 2000$ samples are split into K blocks of L samples, $KL = 2000$. Three different options for (L, K) are considered:

$$(i) \quad L = 10, K = 200; \qquad (ii) \quad L = 100, K = 20; \qquad (iii) \quad L = 2000, K = 1.$$

Assume for the moment $\sigma_s^2 = 4$. The resulting spectrum estimates for $P_x(e^{j\omega})$, in random order, are:



(a) Which resolution is obtained for each choice of (L, K)? Also indicate how this relates to B.

(b) What is the variance on $\hat{P}_B(e^{j\omega})$ in each case? (Make a distinction between outside/inside the passband of $P_s(e^{j\omega})$.)

(c) Which spectrum estimate corresponds to which parameter set (L, K)?

(d) Explain the observed differences between spectrum I and II. In particular, explain why the height of the observed peak in spectrum II is much lower than that in spectrum I.

(e) Explain the observed differences between spectrum II and III. In particular, explain why the height of the observed peak in spectrum III is much lower than that in spectrum II.

(f) Which spectrum estimate would you use to estimate σ_s^2, and how would you estimate it? What is the variance of that estimate?

Solution

(a) Following Hayes, the resolution is $0.89\frac{2\pi}{L}$. Thus,

$$(i) \quad 0.56 > B; \qquad (ii) \quad 0.056 < B; \qquad (iii) \quad 0.0028 < B.$$

(b) The variance is $\frac{1}{K}P_x^2(e^{j\omega})$. Inside the passband, this is $\frac{1}{K}(\sigma_s^2+1)^2$, and outside the passband it is $\frac{1}{K}\cdot 1$.

$$(i) \quad 0.125,\ 0.005; \qquad (ii) \quad 1.25,\ 0.05; \qquad (iii) \quad 25,\ 1.$$

(c) Considering the variance outside the passband, $i \leftrightarrow III$, $ii \leftrightarrow II$, $iii \leftrightarrow I$.

(d) The variance for II is much lower because of the averaging, this explains the lower peak. The resolution for II is coarser. In fact, for II the passband of $P_s(e^{j\omega})$ is only slightly larger than the resolution (about 2 DFT bins), therefore we see only a single spike.

(e) Because K is larger, the variance for III is even lower than for II, making the curve look more smooth. The resolution is very poor (only 10 DFT bins). Because the passband is only a fraction of the DFT bin ($BL/(2\pi) = 0.25$, the source is not resolved), the observed power is reduced by that fraction. Instead of the expected 5, we observe $1 + 4 \cdot 0.25 = 2$. This is the main reason the peak is much lower than for II.

(Note: in the book, there was an example with a single sinusoid in noise, fig 8.5 and fig 8.7. In these examples, the height of the peak in the periodogram scales with L. This is the case where the signal is "not resolved" and falls entirely within the DFT bin, as we have in III. As we go from III to II, L grows, the peak grows. Beyond II, the expected value of the peak doesn't grow, but its variance grows and makes it look large.)

(f) Spectrum estimate I has a large variance, but we can average over the interval; if we use a DFT of length N, there are $BN/(2\pi) = 51$ DFT bins that we can average. The resulting variance of that estimate is

$$\frac{2\pi}{BN}(\sigma_s^2+1)^2$$

Spectrum estimate II has a smaller variance, but we cannot average any bins, the estimate for σ_s^2 is simply the observed peak (minus σ_v^2). Its variance was

$$\frac{(\sigma_s^2+1)^2}{K}$$

Comparing these two, we see

$$\begin{array}{ccc} \frac{2\pi}{BN} & < & \frac{1}{K} \\ 0.0196 & < & 0.0500 \end{array}$$

We prefer to average over the passband of spectrum estimate I.

(To be sure, this difference is only because for spectrum II, L is slightly too large. More accurately, we could take $L = 2\pi/B = 39$ and $K = 51$, in that case we find a single peak that is 1 DFT bin wide, and its variance will be $1/51 = 0.0196$ as in the other case.)

Spectrum estimate III is not useful to estimate σ_s^2. (In detail: we will have to compensate for the fraction of the DFT bin that B occupies. That scaling α enters squared in the expression for the variance. Even if the variance of 1 DFT sample is lower due to the larger averaging, the resulting scaled average is worse by a factor α.)

Delft University of Technology
Faculty of Electrical Engineering, Mathematics, and Computer Science
Circuits and Systems Group

EE4C03 STATISTICAL DIGITAL SIGNAL PROCESSING AND MODELING

4 November 2020, 13:30–16:50
Block 2 (15:20-16:50)
Open book, strictly timed take-home exam. (Electronic) copies of the book and the course slides allowed. No other tools except a basic pocket calculator permitted.
Upload answers during 16:45–16:55

This block consists of three questions (25 points); more than usual, and this will be taken into account during grading. Answer in Dutch or English. Make clear in your answer how you reach the final result; the road to the answer is very important. Write your name and student number on each sheet.

Hint: Avoid losing too much time on detailed calculations, write down the general approach first.

Question 1 (9 points)

Consider a stochastic process consisting of a complex exponential in complex noise. Mathematically, this can be written as

$$x(n) = \alpha e^{j\omega_0 n} + \beta v(n), \tag{1}$$

where α and β are some weighting parameters, ω_0 is the frequency of the complex exponential and $v(n)$ is complex white Gaussian noise with mean 0 and variance 1.

We will fit an all-pole model of order two to this process, which can be written as

$$H(z) = \frac{b(0)}{1 + a(1)z^{-1} + a(2)z^{-2}},$$

with filter coefficients $b(0)$, $a(1)$ and $a(2)$.

Let us first assume that $\alpha = 1$ and $\beta = 0$, i.e., $x(n) = e^{j\omega_0 n}$, and that $N+1$ samples from $x(n)$ are given, i.e., $x(0), x(1), \ldots, x(N)$.

(a) Use the autocorrelation method based on these $N+1$ samples to compute $a(1)$ and $a(2)$. Provide all the steps of the derivation.

Hint: The inverse of the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *is given by* $\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

(b) Use the covariance method based on these $N+1$ samples to compute $a(1)$ and $a(2)$. Provide all the steps of the derivation.

(c) Observe that the solutions for (a) and (b) are different. Which solution do you prefer? Are there any conditions for which the two methods give the same solution?

Let us now assume the general case where both α and β are non-zero.

(d) Compute the correlation sequence of $x(n)$ based on the model (1). Is the process $x(n)$ wide-sense stationary? Why or why not?

(e) Use the Yule Walker model to compute $a(1)$ and $a(2)$. Provide all the steps of the derivation.

(f) If you set $\alpha = 1$ and $\beta = 0$, how does the solution of (e) look like? Does it correspond to the solution of (a) and/or (b) above?

(g) If you now set $\alpha = 0$ and $\beta = 1$, how does the solution of (e) look like? Can you interpret that solution? Or in other words, is this the solution you would expect?

Solution

(a) For the autocorrelation method, we can directly make use of the (deterministic) correlation function

$$r_x(k) = \sum_{n=k}^{N} x(n)x^*(n-k) = \sum_{n=k}^{N} e^{j\omega_0 n} e^{-j\omega_0(n-k)} = (N+1-k)e^{j\omega_0 k}.$$

The related normal equations then look like

$$\begin{bmatrix} r_x(0) & r_x^*(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = - \begin{bmatrix} r_x(1) \\ r_x(2) \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} N+1 & Ne^{-j\omega_0} \\ Ne^{j\omega_0} & N+1 \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = - \begin{bmatrix} Ne^{j\omega_0} \\ (N-1)e^{j2\omega_0} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = - \frac{1}{(N+1)^2 - N^2} \begin{bmatrix} N+1 & -Ne^{-j\omega_0} \\ -Ne^{j\omega_0} & N+1 \end{bmatrix} \begin{bmatrix} Ne^{j\omega_0} \\ (N-1)e^{j2\omega_0} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = - \frac{1}{2N+1} \begin{bmatrix} 2Ne^{j\omega_0} \\ -e^{j2\omega_0} \end{bmatrix}$$

(b) For the covariance method, we need to use a different type of (deterministic) correlation function

$$r_x(k,l) = \sum_{n=2}^{N} x(n-l)x^*(n-k) = \sum_{n=2}^{N} e^{j\omega_0(n-l)} e^{-j\omega_0(n-k)} = (N-1)e^{j\omega_0(k-l)}.$$

The normal equations are then given by

$$\begin{bmatrix} r_x(1,1) & r_x(1,2) \\ r_x(2,1) & r_x(2,2) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = - \begin{bmatrix} r_x(1,0) \\ r_x(2,0) \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & e^{-j\omega_0} \\ e^{j\omega_0} & 1 \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = - \begin{bmatrix} e^{j\omega_0} \\ e^{j2\omega_0} \end{bmatrix}.$$

These two equations are actually the same, which means we have multiple solutions. One such solution is given by $a(1) = -e^{j\omega_0}$ and $a(2) = 0$. Another example solution is $a(1) = -0.5e^{j\omega_0}$ and $a(2) = -0.5e^{j2\omega_0}$.

(c) The covariance method can be preferred since it gives a zero prediction error. If $N \to \infty$ the solution of the autocorrelation method is also a solution for the covariance method.

(d) The statistical correlation function of $x(n)$ is computed as

$$r_x(k) = E\{x(n)x^*(n-k)\} = \alpha^2 e^{j\omega_0 k} + \beta^2 \delta(k).$$

Since the mean is always zero and the correlation function only depends on the time difference, the process is stationary.

(e) The Yule-Walker model is given by

$$\begin{bmatrix} r_x(0) & r_x^*(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = - \begin{bmatrix} r_x(1) \\ r_x(2) \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \alpha^2+\beta^2 & \alpha^2 e^{-j\omega_0} \\ \alpha^2 e^{j\omega_0} & \alpha^2+\beta^2 \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = - \begin{bmatrix} \alpha^2 e^{j\omega_0} \\ \alpha^2 e^{j2\omega_0} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = -\frac{1}{(\alpha^2+\beta^2)^2-\alpha^4} \begin{bmatrix} \alpha^2+\beta^2 & -\alpha^2 e^{-j\omega_0} \\ -\alpha^2 e^{j\omega_0} & \alpha^2+\beta^2 \end{bmatrix} \begin{bmatrix} \alpha^2 e^{j\omega_0} \\ \alpha^2 e^{j2\omega_0} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = -\frac{1}{(\alpha^2+\beta^2)^2-\alpha^4} \begin{bmatrix} \alpha^2\beta^2 e^{j\omega_0} \\ \alpha^2\beta^2 e^{j2\omega_0} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} a(1) \\ a(2) \end{bmatrix} = -\frac{1}{2+\beta^2/\alpha^2} \begin{bmatrix} e^{j\omega_0} \\ e^{j2\omega_0} \end{bmatrix}$$

(f) We then obtain the solution $a(1) = -0.5e^{j\omega_0}$ and $a(2) = -0.5e^{j2\omega_0}$. Actually, in that case, multiple solutions are again possible since the Yule-Walker method is then the same as the covariance method, i.e., the method in (b).

(g) We then obtain the solution $a(1) = 0$ and $a(2) = 0$. This makes sense since we cannot predict white noise.

Question 2 (8 points)

In this question, we will develop the concept of optimal filtering but applied to compressed signals. Assume all signals are real-valued. Let us consider a stationary stochastic process $x(n)$ which depends on some stationary desired signal $d(n)$. Further assume the auto-correlation sequence of $x(n)$ is given by $r_x(k) = E\{x(n)x(n-k)\}$ and the cross-correlation sequence between $d(n)$ and $x(n)$ by $r_{dx}(k) = E\{d(n)x(n-k)\}$.

Let us now assume the signal $x(n)$ is compressed using the following steps: i) $x(n)$ is split into non-overlapping windows of size N, which we will denote as $\mathbf{x}_i = [x(iN), x(iN+1), \ldots, x(iN+N-1)]^T$; ii) every window $\mathbf{x}_i$ is compressed by applying an $M \times N$ matrix $\mathbf{C}$ to it ($M < N$), leading to a smaller window $\mathbf{y}_i = \mathbf{C}\mathbf{x}_i$; iii) the smaller windows $\mathbf{y}_i$ are concatenated leading to the sequence $y(m)$ which is a compressed version of $x(n)$ (we use the index m instead of n here because it is related to a lower sampling rate). In short, every N samples of $x(n)$ are transformed into $M < N$ samples of $y(m)$ using the wide matrix $\mathbf{C}$.

(a) Express the auto-correlation matrix $\mathbf{R}_x = E\{\mathbf{x}_i\mathbf{x}_i^T\}$ as a function of the autocorrelation sequence $r_x(k)$.

Hint: Focus on a single entry of the matrix and compute the related correlation.

(b) If we define the windowed version of $d(n)$ as $\mathbf{d}_i = [d(iN), d(iN+1), \dots, d(iN+N-1)]^T$, express the cross-correlation matrix $\mathbf{R}_{dx} = E\{\mathbf{d}_i \mathbf{x}_i^T\}$ as a function of the cross-correlation sequence $r_{dx}(k)$.

Hint: Use the same hint as in (a).

(c) Express the auto-correlation matrix $\mathbf{R}_y = E\{\mathbf{y}_i \mathbf{y}_i^T\}$ as a function of $\mathbf{R}_x$ from (a). Similarly, express the cross-correlation matrix $\mathbf{R}_{dy} = E\{\mathbf{d}_i \mathbf{y}_i^T\}$ as a function of $\mathbf{R}_{dx}$ from (b).

(d) Suppose we want to estimate $\mathbf{d}_i$ from $\mathbf{y}_i$ using an $M \times N$ filter $\mathbf{W}$ such that $\hat{\mathbf{d}}_i = \mathbf{W}^T \mathbf{x}_i$. Find the optimal filter $\mathbf{W}$ by solving

$$\min_{\mathbf{W}} E\{\|\mathbf{W}^T \mathbf{y}_i - \mathbf{d}_i\|^2\}.$$

Give all the steps of the derivation and express the solution as a function of $\mathbf{R}_x$ and $\mathbf{R}_{dx}$ using the results derived in (c).

Hint: Solve the problem for the kth filter $\mathbf{w}_k$ (kth column of $\mathbf{W}$) which is used to estimate $d(iN+k-1)$ (kth entry of $\mathbf{d}_i$). Then stack all the solutions.

(e) Suppose now that $x(n) = d(n)$ and both are white Gaussian noise with mean 0 and variance 1. How can then $\hat{\mathbf{d}}_i$ be expressed as a function of $\mathbf{d}_i$? Is $\hat{\mathbf{d}}_i = \mathbf{d}_i$? Why or why not?

(f) Based on your derivation in (d), express the steepest gradient descent recursion to compute $\mathbf{W}$. From that recursion, derive the LMS algorithm to adaptively compute $\mathbf{W}$ using knowledge of $\mathbf{y}_i$ and $\mathbf{d}_i$.

Solution

(a) Looking at the entry at row k and column l, we can derive

$$[\mathbf{R}_x]_{k,l} = E\{x(iN+k-1)x(iN+l-1)\} = r_x(k-l).$$

So the full auto-correlation matrix looks like

$$\mathbf{R}_x = \begin{bmatrix} r_x(0) & r_x(1) & \dots & r_x(N-1) \\ r_x(1) & r_x(0) & \dots & r_x(N-2) \\ \vdots & \vdots & & \vdots \\ r_x(N-1) & r_x(N-2) & \dots & r_x(0) \end{bmatrix}.$$

(b) Looking at the entry at row k and column l, we can derive

$$[\mathbf{R}_{dx}]_{k,l} = E\{d(iN+k-1)x(iN+l-1)\} = r_{dx}(k-l).$$

So the full cross-correlation matrix looks like

$$\mathbf{R}_{dx} = \begin{bmatrix} r_{dx}(0) & r_{dx}(1) & \dots & r_{dx}(N-1) \\ r_{dx}(1) & r_{dx}(0) & \dots & r_{dx}(N-2) \\ \vdots & \vdots & & \vdots \\ r_{dx}(N-1) & r_{dx}(N-2) & \dots & r_{dx}(0) \end{bmatrix}.$$

(c) Using the fact that $\mathbf{y}_i = \mathbf{C}\mathbf{x}_i$, we can express $\mathbf{R}_y$ and $\mathbf{R}_{dy}$ as

$$\mathbf{R}_y = \mathbf{C}\mathbf{R}_x\mathbf{C}^T,$$
$$\mathbf{R}_{dy} = \mathbf{R}_{dx}\mathbf{C}^T$$

(d) The cost for the filter $\mathbf{w}_k$ is given by

$$J(\mathbf{w}_k) = \mathbf{w}_k^T \mathbf{R}_y \mathbf{w}_k - \mathbf{w}_k^T [\mathbf{R}_{dy}^T]_{:,k+1} + E\{d^2(iN + k - 1)\}.$$

Taking the derivative w.r.t. $\mathbf{w}_k$, and setting it to zero, we obtain

$$\mathbf{R}_y \mathbf{w}_k = [\mathbf{R}_{dy}^T]_{:,k+1},$$

and hence the solution for $\mathbf{w}_k$ is

$$\mathbf{w}_k = \mathbf{R}_y^{-1} [\mathbf{R}_{dy}^T]_{:,k+1}.$$

Stacking all the filters, we obtain

$$\mathbf{W} = \mathbf{R}_y^{-1} \mathbf{R}_{dy}^T = (\mathbf{C}\mathbf{R}_x\mathbf{C}^T)^{-1} \mathbf{C}\mathbf{R}_{dx}^T.$$

(e) In that case, we obtain

$$\mathbf{W} = (\mathbf{C}\mathbf{R}_x\mathbf{C}^T)^{-1} \mathbf{C}\mathbf{R}_x^T = (\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{C}.$$

As a result, the estimate for $\mathbf{d}_i$ is given by

$$\hat{\mathbf{d}}_i = \mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1} \mathbf{y}_i = \mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1} \mathbf{C}\mathbf{d}_i.$$

In general, $\hat{\mathbf{d}}_i$ is different from $\mathbf{d}_i$. Only when $\mathbf{C}$ is square invertible, they are the same.

(f) The steepest gradient descent recursion for $\mathbf{W}$ is given by

$$\mathbf{W}^{(i)} = \mathbf{W}^{(i-1)} - \mu(\mathbf{R}_y \mathbf{W}^{(i-1)} - \mathbf{R}_{dy}^T).$$

Dropping the expected values in $\mathbf{R}_y$ and $\mathbf{R}_{dy}$ we obtain

$$\begin{aligned} \hat{\mathbf{W}}^{(i)} &= \hat{\mathbf{W}}^{(i-1)} - \mu(\mathbf{y}_i \mathbf{y}_i^T \hat{\mathbf{W}}^{(i-1)} - \mathbf{y}_i \mathbf{d}_i^T) \\ &= \hat{\mathbf{W}}^{(i-1)} - \mu \mathbf{y}_i (\mathbf{y}_i^T \hat{\mathbf{W}}^{(i-1)} - \mathbf{d}_i^T) \\ &= \hat{\mathbf{W}}^{(i-1)} - \mu \mathbf{y}_i \mathbf{e}_i^T, \end{aligned}$$

where the error vector is given by

$$\mathbf{e}_i = \hat{\mathbf{W}}^{(i-1)} \mathbf{y}_i - \mathbf{d}_i.$$

Question 3 (8 points)

Suppose we want to use spectral estimation techniques for localizing an acoustic source that is picked up by a number of microphones. Let us for simplicity assume that positions in 3D are indexed by some parameter $\phi \in \mathbb{N}$. The source is at position ϕ_0 and transmits a stationary white Gaussian signal $s(n)$ with mean 0 and variance 1. Microphone k, with $k = 1, 2, \ldots, K$, is at position ϕ_k and picks up the signal

$$x_k(n) = a_k(\phi_0) s(n) + v_k(n).$$

Here, $a_k(\phi)$ is the attenuation due to the propagation from a source at position ϕ to the position ϕ_k of microphone k. Further, $v_k(n)$ is stationary white Gaussian noise with mean 0 and variance σ^2 and the noises on the different microphones are assumed mutually independent. We also define the K-dimensional column vectors $\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_K(n)]^T$, $\mathbf{v}(n) = [v_1(n), v_2(n), \ldots, v_K(n)]^T$, and $\mathbf{a}(\phi) = [a_1(\phi), a_2(\phi), \ldots, a_K(\phi)]^T$. Although this might not be true in practice, we assume here for simplicity that $\mathbf{a}(\phi)$ is normalized for every position ϕ, i.e., $\|\mathbf{a}(\phi)\| = 1, \ \forall \phi$.

(a) Express $\mathbf{x}(n)$ as a function of $\mathbf{v}(n)$, $\mathbf{a}(\phi_0)$, and $s(n)$. Based on this model, compute the correlation matrix of $\mathbf{x}(n)$, denoted by $\mathbf{R}_x = E\{\mathbf{x}(n)\mathbf{x}^T(n)\}$.

(b) If the noise variance $\sigma^2 = 0$, then what is the rank of $\mathbf{R}_x$ and why?

Suppose now that we want to estimate the position ϕ_0 of the source using techniques that are related to spectral estimation. Therefore, we first consider an arbitrary position ϕ and we apply the *position-dependent* filter $\mathbf{w}$ to $\mathbf{x}(n)$ leading to the output $y(n) = \mathbf{w}^T\mathbf{x}(n)$. Then we compute the output power of $y(n)$ resulting in $P(\phi) = \mathbf{w}^T\mathbf{R}_x\mathbf{w}$, which can be interpreted as the "position" spectrum. Finally, we estimate the source position ϕ_0 by selecting the value of ϕ for which $P(\phi)$ is maximized.

(c) Give the expression for the filter $\mathbf{w}$ related to the periodogram. Using the model derived in (a), give the expression for the related position spectrum $P(\phi)$. What is the maximum of this spectrum and for which value of ϕ is it obtained?

(d) Give the expression for the filter $\mathbf{w}$ related to the minimum variance spectral estimation method. Using the model derived in (a), give the expression for the related position spectrum $P(\phi)$. What is the maximum of this spectrum and for which value of ϕ is it obtained?

Hint: You will have to use a specific form of the matrix inversion lemma, i.e., $(c + \mathbf{x}\mathbf{x}^T)^{-1} = c^{-1}\mathbf{I} - c^{-1}(c + \mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}\mathbf{x}^T$.

For finding the filter related to the MUSIC method, we solve the following problem. For every source position ϕ we minimize the distance between $\mathbf{w}$ and $\mathbf{a}(\phi)$ subject to the fact that $\mathbf{w}$ is orthogonal to the noise subspace of $\mathbf{R}_x$ described by the $K \times (K - r)$ matrix $\mathbf{U}_n$ [note that r is the rank you have to derive in (b)]. This problem can be mathematically formulated as

$$\min_{\mathbf{w}} \|\mathbf{w} - \mathbf{a}(\phi)\|^2 \quad \text{subject to} \quad \mathbf{U}_n^T\mathbf{w} = \mathbf{0}. \tag{2}$$

(e) Solve problem (2) using the method of Lagrange multipliers. Note that you have $K - r$ linear constraints so your Lagrange multiplier is a vector $\boldsymbol{\lambda}$ of size $K - r$. Prove that the solution is given by $\mathbf{w} = (\mathbf{I} - \mathbf{U}_n\mathbf{U}_n^T)\mathbf{a}(\phi)$.

(f) The useful signal energy of this MUSIC filter in case there is a source at position ϕ is given by $|\mathbf{w}^T\mathbf{a}(\phi)|^2$. Show that maximizing this useful signal energy is the same as maximizing the so-called MUSIC pseudo-spectrum given by $[\mathbf{a}^T(\phi)\mathbf{U}_n\mathbf{U}_n^T\mathbf{a}(\phi)]^{-1}$.

Solution

(a) Stacking the results for the different microphones, we obtain

$$\mathbf{x}(n) = \mathbf{a}(\phi_0)s(n) + \mathbf{v}(n).$$

The correlation matrix is given by

$$\mathbf{R}_x = \mathbf{a}(\phi_0)\mathbf{a}^T(\phi_0) + \sigma^2\mathbf{I}.$$

(b) From the above expression we see that when $\sigma^2 = 0$, every column (or row) of $\mathbf{R}_x$ is up to a scaling the same. So the rank, or the number of independent columns (rows) is 1.

(c) In the periodogram, the applied filter is matched to the waveform we are looking for. In classical spectrum estimation, this is the complex exponential with a varying frequency, so here it is simply given by

$$\mathbf{w} = \mathbf{a}(\phi).$$

The position spectrum is then given by

$$P(\phi) = \mathbf{w}^T \mathbf{R}_x \mathbf{w} = \mathbf{a}^T(\phi)\mathbf{R}_x \mathbf{a}(\phi) = |\mathbf{a}^T(\phi)\mathbf{a}(\phi_0)|^2 + \sigma^2.$$

When ϕ is matched to ϕ_0, the maximum value is obtained and this is given by $1 + \sigma^2$.

(d) From spectral estimation, we know that the minimum variance filter is given by

$$\mathbf{w} = \frac{\mathbf{R}_x^{-1}\mathbf{a}(\phi)}{\mathbf{a}^T(\phi)\mathbf{R}_x^{-1}\mathbf{a}(\phi)}.$$

Using the matrix inversion lemma, the related position spectrum can be written as

$$\begin{aligned} P(\phi) = \mathbf{w}^T \mathbf{R}_x \mathbf{w} &= \frac{1}{\mathbf{a}^T(\phi)\mathbf{R}_x^{-1}\mathbf{a}(\phi)} \\ &= \frac{\sigma^2(1+\sigma^2)}{\mathbf{a}^T(\phi)[(1+\sigma^2)\mathbf{I} - \mathbf{a}(\phi_0)\mathbf{a}^T(\phi_0)]\mathbf{a}(\phi)} \\ &= \frac{\sigma^2(1+\sigma^2)}{1+\sigma^2 - |\mathbf{a}^T(\phi)\mathbf{a}(\phi_0)|^2} \end{aligned}$$

Again we see than when ϕ is matched to ϕ_0, the maximum value is obtained and this is again given by $1 + \sigma^2$.

(e) The Lagrangian of the problem is given by

$$\begin{aligned} L(\mathbf{w}) &= \|\mathbf{w} - \mathbf{a}(\phi)\|^2 + \mathbf{w}^T \mathbf{U}_n \boldsymbol{\lambda} \\ &= \mathbf{w}^T\mathbf{w} + \mathbf{a}^T(\phi)\mathbf{a}(\phi) - 2\mathbf{w}^T\mathbf{a}(\phi) + \mathbf{w}^T\mathbf{U}_n\boldsymbol{\lambda}. \end{aligned}$$

Taking the derivative to $\mathbf{w}$ and setting it to zero, we obtain

$$2\mathbf{w} - 2\mathbf{a}(\phi) + \mathbf{U}_n\boldsymbol{\lambda} = \mathbf{0}.$$

Solving this for $\mathbf{w}$ leads to

$$\mathbf{w} = \mathbf{a}(\phi) - \frac{1}{2}\mathbf{U}_n\boldsymbol{\lambda}.$$

Plugging this into the constraint, we get

$$\boldsymbol{\lambda} = 2\mathbf{U}_n^T\mathbf{a}(\phi).$$

So the solution for $\mathbf{w}$ finally is

$$\mathbf{w} = (\mathbf{I} - \mathbf{U}_n\mathbf{U}_n^T)\mathbf{a}(\phi).$$

(f) The useful signal energy of the MUSIC filter if there is a source at position ϕ is given by

$$|\mathbf{w}^T\mathbf{a}(\phi)|^2 = |\mathbf{a}^T(\phi)(\mathbf{I} - \mathbf{U}_n\mathbf{U}_n^T)\mathbf{a}(\phi)|^2 = |1 - \mathbf{a}^T(\phi)\mathbf{U}_n\mathbf{U}_n^T\mathbf{a}(\phi)|^2.$$

Since the second term $\mathbf{a}(\phi)\mathbf{U}_n\mathbf{U}_n^T\mathbf{a}(\phi)$ is positive and smaller than 1, the useful energy is maximal if that term is minimal. And since that term is in the denominator of the MUSIC spectrum it means that the useful energy is maximal if the MUSIC spectrum is maximal.