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

EE4C03 STATISTICAL DIGITAL SIGNAL PROCESSING AND MODELING

23 January 2020, 18:30–21:30

Open book exam: copies of the book by Hayes and the course slides allowed.

This exam has four questions (40 points)

Question 1 (10 points)

Assume that $s(n)$ is a stationary random process with zero mean and autocorrelation function $r_s(k) = \delta(k)$. We now form a random process $x(n)$ as follows:

$$x(n) = s(n) + f(n),$$

where $f(n)$ is a known *deterministic* sequence.

Subsequently, we filter the sequence $x(n)$ with $h(n)$, i.e., $y(n) = x(n) * h(n)$, where the system function is given by

$$H(z) = \frac{1}{1 - \frac{1}{2}z^{-1}}.$$

(a) Compute the mean and the autocorrelation of $x(n)$.

(b) Is $x(n)$ stationary for every function $f(n)$? If not, can you give a counter example?

Are there functions $f(n)$ for which $x(n)$ is stationary? If so, what are the conditions on $f(n)$ for $x(n)$ to be stationary?

Assume from now on that $f(n) = f$ is a constant.

(c) Compute the power spectrum, $P_x(e^{j\omega})$, of $x(n)$. Make a sketch of this power spectrum.

(d) What is the power spectrum, $P_y(e^{j\omega})$, of $y(n)$? Make a sketch of this power spectrum.

(e) What is the cross-power spectrum, $P_{xy}(e^{j\omega})$, between $x(n)$ and $y(n)$?

Solution

(a) Taking into account that $f(n)$ is deterministic, the mean of $x(n)$ is given by

$$m_x(n) = E\{x(n)\} = E\{s(n)\} + f(n) = f(n),$$

(note that it is time-varying i.e. not stationary), and the autocorrelation is

$$\begin{aligned} r_x(k,l) &= E\{x(k)x(l)\} = E\{[s(k) + f(k)][s(l) + f(l)]\} \\ &= E\{s(k)s(l)\}f(k)f(l) = \delta(k-l) + f(k)f(l). \end{aligned}$$

(b) Clearly, $x(n)$ is not stationary in general. Take for instance the deterministic function $f(n) = n$, for which $m_x(n)$ is not independent of n. However, there are functions $f(n)$ for which $x(n)$ is stationary. Since $m_x(n)$ has to be independent of n for $x(n)$ to be stationary, the function $f(n)$ has to be constant, i.e., $f(n) = f$. If we plug this in $r_x(k, l)$, we obtain $r_x(k, l) = \delta(k - l) + f^2 = r_x(k - l)$, which only depends on $k - l$. As a result, the function $f(n)$ has to be constant for $x(n)$ to be stationary.

(c) The power spectrum for a stationary $x(n)$ is given by

$$P_x(e^{j\omega}) = 1 + 2\pi f^2 \delta(\omega).$$

Note that this function is only valid for $-\pi \leq \omega \leq \pi$ (beyond that the spectrum is periodic).

(d) The power spectrum of $y(n)$ is

$$\begin{aligned} P_y(e^{j\omega}) &= P_x(e^{j\omega})|H(e^{j\omega})|^2 \\ &= P_x(e^{j\omega})\frac{1}{\frac{5}{4} - \cos\omega} \\ &= \frac{1}{\frac{5}{4} - \cos\omega} + 8\pi f^2 \delta(\omega). \end{aligned}$$

(e) The cross-power spectrum between $x(n)$ and $y(n)$ is

$$\begin{aligned} P_{xy}(e^{j\omega}) &= P_x(e^{j\omega})H(e^{-j\omega}) \\ &= P_x(e^{j\omega})\frac{1}{1 - \frac{1}{2}e^{j\omega}} \\ &= \frac{1}{1 - \frac{1}{2}e^{j\omega}} + 4\pi f^2 \delta(\omega). \end{aligned}$$

Question 2 (12 points)

Suppose we have a random process $x(n)$ with zero mean and *complex-valued* autocorrelation $r_x(k)$ for which we would like to obtain an all-pole model of the form

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

(a) Write down the Yule-Walker equations, which express $b(0)$, $a(1)$, and $a(2)$ in terms of $r_x(k)$.

Hint: Don't forget the equation for $b(0)$, and remember that $r_x(k)$ is complex-valued.

(b) For the optimal filter, explain why the denominator of $H(z)$ can be interpreted as a prediction error filter.

(c) What is the orthogonality principle?

Use the orthogonality principle to derive the mean square of the related prediction error, denoted as ϵ.

Suppose now that we have reasons to believe that the signal is periodic and, consequently, the poles of the model should lie on the unit circle. Assume that for our second-order model, we have two poles $e^{j\theta}$ and $e^{-j\theta}$.

(d) Express $a(1)$ and $a(2)$ as a function of θ.

(e) Use the result of (d) and the result of (a) to find relations between θ and $r_x(k)$.

From these relations, derive conditions on $r_x(k)$ such that solving the Yule-Walker equations results in a model with the two mentioned poles.

In a different context, assume now that $r_x(0) = 2$, $r_x(1) = 0.5(1 + j)$, and $r_x(2) = 0.5j$.

(f) Use the Levinson-Durbin recursion to solve the Yule-Walker equations of (a) explicitly.

What are the reflection coefficients?

Is this a stable filter or not (explain why)?

Solution

(a) The Yule-Walker equations are given by

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

Further, we have

$$|b(0)|^2 = r_x(0) + a(1)r_x^*(1) + a(2)r_x^*(2),$$

so we could for instance take $b(0)$ real-valued and equal to

$$b(0) = |r_x(0) + a(1)r_x^*(1) + a(2)r_x^*(2)|^{1/2}.$$

(b) The optimal filter basically minimizes $E\{|e(n)|^2\}$, where $E(z) = A(z)X(z)$ with $A(z) = 1 + a(1)z^{-1} + a(2)z^{-2}$. So we can express $E(z)$ in the time domain as

$$\begin{aligned} e(n) &= x(n) + a(1)x(n-1) + a(2)x(n-2) \\ &= x(n) - [-a(1)x(n-1) - a(2)x(n-2)], \end{aligned}$$

which is the error we make when we predict $x(n)$ using past data as $-a(1)x(n-1) - a(2)x(n-2)$. And hence, we can view $E(z)$ as the prediction error and $A(z)$ as a filter that minimizes this prediction error.

(c) The orthogonality principle states that the error $e(n)$ is orthogonal to the input of the different filter taps, i.e., $x(n-k)$ for $k = 1, 2$. This means that $E\{e(n)x^*(n-1)\} = E\{e(n)x^*(n-2)\} = 0$. The mean square prediction error can then be written as

$$\begin{aligned} \epsilon &= E\{|e(n)|^2\} = E\{e(n)e^*(n)\} \\ &= E\{e(n)x^*(n)\} + a(1)E\{e(n)x^*(n-1)\} + a(2)E\{e(n)x^*(n-2)\}. \end{aligned}$$

Since the last two terms are zero according to the orthogonality principle, we obtain

$$\epsilon = E\{e(n)x^*(n)\} = r_x(0) + a(1)r_x^*(1) + a(2)r_x^*(2).$$

(d) It is clear that for this scenario, we have $A(z) = (1 - e^{j\theta}z^{-1})(1 - e^{-j\theta}z^{-1})$, which can be re-written as $A(z) = 1 - 2\cos\theta z^{-1} + z^{-2}$. This means that $a(1) = -2\cos\theta$ and $a(2) = 1$.

(e) It is easy to see that the first Yule-Walker equation gives us

$$-2\cos\theta r_x(0) + r_x^*(1) = -r_x(1) \quad \Rightarrow \quad \cos\theta = \frac{\Re\{r_x(1)\}}{r_x(0)},$$

whereas the second equation results into

$$-2\cos\theta r_x(1) + r_x(0) = -r_x(2) \qquad \Rightarrow \quad \cos\theta = \frac{r_x(0) + r_x(2)}{2r_x(1)}.$$

So the relation we need between $r_x(0)$, $r_x(1)$ and $r_x(2)$ is

$$-1 \leq \frac{\Re\{r_x(1)\}}{r_x(0)} = \frac{r_x(0) + r_x(2)}{2r_x(1)} \leq 1.$$

(f) The first-order model gives us

$$\begin{aligned}
\Gamma_1 &= -\frac{r_x(1)}{r_x(0)} = -\frac{1+j}{4}, \\
\epsilon_1 &= r_x(0)(1 - |\Gamma_1|^2) = 2(1 - \frac{1}{8}) = \frac{7}{4}, \\
\mathbf{a}_1 &= \begin{bmatrix} 1 \\ \Gamma_1 \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{1+j}{4} \end{bmatrix}.
\end{aligned}$$

For the second-order model, we obtain

$$\begin{aligned}
\gamma_1 &= r_x(2) + a_1(1)r_x(1) = \frac{j}{2} - \frac{1+j}{4}\frac{1+j}{2} = \frac{j}{4}, \\
\Gamma_2 &= -\frac{\gamma_1}{\epsilon_1} = -\frac{j}{4}\frac{4}{7} = -\frac{j}{7}, \\
\epsilon_2 &= \epsilon_1(1 - |\Gamma_2|^2) = \frac{7}{4}(1 - \frac{1}{49}) = \frac{12}{7}, \\
\mathbf{a}_2 &= \begin{bmatrix} 1 \\ -\frac{1+j}{4} \\ 0 \end{bmatrix} - \frac{j}{7}\begin{bmatrix} 0 \\ -\frac{1-j}{4} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{3(1+j)}{14} \\ -\frac{j}{7} \end{bmatrix}.
\end{aligned}$$

We see that the filter is stable since the reflection coefficients are smaller than 1 in absolute value.

Question 3 (8 points)

Suppose we have a continuous-time random process of the form

$$x(t) = e^{j2\pi f_0 t} + w(t),$$

where $f_0 = 2$ kHz and $w(t)$ is a zero mean process with autocorrelation function $r_w(\tau) = 0.1\delta(\tau)$. Only 5 seconds of the signal have been recorded and are available for processing. We would now like to estimate the power spectrum of $x(t)$ at a resolution of at least 10 Hz.

(a) Suppose we want to estimate the power spectrum at a bandwidth of 2.5 kHz using Bartlett's method of periodogram averaging. What is the minimum sampling rate we require for that, and what is the minimum data segment length (expressed in number of samples) to obtain the desired resolution?

(b) Using the minimum segment length of (a), and 5 seconds of measurements of the signal, how many segments are available for averaging?

(c) How does the sampling rate affect the resolution and variance of your estimate? Are there any benefits to sampling at a higher rate than the one computed in (a)?

(d) Given only the estimated averaged periodogram $\hat{P}_x(e^{j\omega})$, explain how you can estimate the frequency f_0 using MUSIC.

Solution

(a) The sampling rate we need to reconstruct a (single-sided) bandwidth of 2.5 kHz is $f_s = 5$ kHz, and to get a resolution of $\Delta f = 10$ Hz (in analog frequency) means we want a resolution (in radians) of

$$\Delta\omega = 2\pi\frac{\Delta f}{f_s} = 4\pi \times 10^{-3}.$$

Since the resolution of Bartlett's method of periodogram averaging for a section length of L samples is given by

$$\Delta\omega = 0.89\frac{2\pi}{L},$$

we need a section length of at least

$$\left\lceil 0.89\frac{2\pi}{\Delta\omega} \right\rceil = 445 \text{ samples.}$$

(b) Sampling at 5 kHz, 5 seconds of data corresponds to $N = 25,000$ samples. So we have $K = \lfloor N/L \rfloor = 56$ sections available for averaging.

(c) If the sampling rate increases, the resolution $\Delta\omega$ decreases and the section length L increases. But if the sampling rate increases, also the total number of samples N will increase, and thus the ratio $K = N/L$ will remain the same. Since K determines the variance of the averaged periodogram, increasing the sampling rate has little effect, as long as the sampling rate is above the Nyquist rate (4 kHz in this case).

(d) From $\hat{P}_x(e^{j\omega})$, we use the inverse discrete-time Fourier transform to compute the estimated autocorrelation function $\hat{r}_x(k)$. From this function, we build the $p \times p$ correlation matrix (p should be at least two)

$$\hat{\mathbf{R}}_x = \begin{bmatrix} \hat{r}_x(0) & \hat{r}_x^*(1) & \cdots & \hat{r}_x^*(p) \\ \hat{r}_x(1) & \hat{r}_x(0) & \cdots & \hat{r}_x^*(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_x(p) & \hat{r}_x(p-1) & \cdots & \hat{r}_x(0) \end{bmatrix}.$$

We compute the eigenvalue decomposition of $\hat{\mathbf{R}}_x$ as

$$\hat{\mathbf{R}}_x = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^H \\ \mathbf{v}_2^H \\ \cdots \\ \mathbf{v}_p^H \end{bmatrix},$$

and since we know that we are looking for a single frequency, the noise subspace is spanned by the matrix $\mathbf{V}_n = [\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_p]$. The frequency $\omega_0 = 2\pi f_0$ is finally found by solving

$$\hat{\omega}_0 = \arg_\omega \max \frac{1}{\mathbf{e}(\omega)^H \mathbf{V}_n \mathbf{V}_n^H \mathbf{e}(\omega)},$$

where $\mathbf{e}(\omega) = [1, e^{j\omega}, e^{j2\omega}, \ldots, e^{j(p-1)\omega}]^T$.

Question 4 (10 points)

Consider a radar system where the distance from the radar to the target is given by x. Assume the target distance x changes with time n, and can thus also be represented by the function $x(n)$. Further, assume the radar system can measure the distance $x(n)$ through a noisy measurement $y(n)$, or in other words, we have

$$y(n) = x(n) + v(n),$$

where $v(n)$ is additive white Gaussian measurement noise with zero mean and variance σ_v^2.

(a) Assume we are at time step n and we have received all measurements from time step 0 up to n, i.e., $y(0), y(1), \ldots, y(n)$. Further, assume we have no knowledge of how the target moves, and hence we simply assume $x(n)$ is a constant x. Then, compute the least squares solution for the target distance x using all the measurements from time step 0 to n. What does this solution represent?

Hint: View x as a filter that filters an input signal that is equal to the constant function 1.

(b) In order to write the solution of the above problem in a recursive way, derive the RLS update equations for $x(n)$. Simplify the equations as much as possible.

What is a good initial condition for the parameters?

(c) Derive the LMS updating rule for $x(n)$.

(d) Is there a step size μ to make this LMS method equivalent to the RLS method? If so, is this step size constant or changing in time?

Hint: First write the RLS update equation in a similar form as the LMS recursion.

(e) Now, assume that we know how the distance of the target changes in time n, and assume we model this as

$$x(n+1) = 2x(n) + w(n),$$

where $w(n)$ is additive white Gaussian process noise with zero mean and variance σ_w^2. Based on this model, derive the Kalman update equations to estimate the target distance $x(n)$.

Will this work better than the other schemes (and why)?

Solution

(a) Let us define

$$\mathbf{1}_n = [1, 1, \ldots, 1] \in \mathbb{R}^{1\times(n+1)}, \quad \mathbf{y}_n = [y(0), y(1), \ldots, y(n)] \in \mathbb{R}^{1\times(n+1)}.$$

Then we can write the data model for a constant $x(n)$ as

$$\mathbf{y}_n = \mathbf{1}_n x + \mathbf{v}_n,$$

and the least squares solution for x is given by

$$\hat{x}(n) = (\mathbf{1}_n \mathbf{1}_n^T)^{-1} \mathbf{1}_n \mathbf{y}_n = \frac{1}{n+1} \sum_{i=0}^{n} y(n),$$

which basically is a simple average of the $n+1$ measurements from time step 0 up to n.

(b) The RLS updating rules are given by

$$\begin{aligned} P_{n+1} &= P_n - \frac{P_n^2}{1+P_n}, \\ \theta_{n+1} &= \theta_n + y(n+1), \\ \hat{x}(n+1) &= P_{n+1}\theta_{n+1}. \end{aligned}$$

Note that an updating rule for P_n is not really required since we know from (a) that $P_n = 1/(n+1)$ so we can also directly compute this in every time step. A good initial condition is $P_0 = 1$ and $\theta_0 = y(0)$.

(c) The LMS updating rule is

$$\hat{x}(n+1) = \hat{x}(n) - \mu 1(1\hat{x}(n) - y(n)) = (1-\mu)\hat{x}(n) + \mu y(n).$$

(d) From the fact that the least squares solution is nothing but a simple average, we can also write RLS as

$$\hat{x}(n+1) = \frac{n}{n+1}\hat{x}(n) + \frac{1}{n+1}y(n+1).$$

Comparing the two equations, we see that LMS is running behind one sample compared to RLS since $\hat{x}(n+1)$ is using $y(n)$ for LMS but $y(n+1)$ for RLS (this is due to the definition of LMS in the book). However, we see that $\hat{x}(n+1)$ from LMS will be equal to $\hat{x}(n)$ from RLS if we take $\mu = 1/n$ so μ should be made varying with time.

(e) The Kalman update equations are

$$\begin{aligned} \hat{x}(n|n-1) &= 2\hat{x}(n-1|n-1), \\ P(n|n-1) &= 4P(n-1|n-1) + \sigma_w^2, \\ K(n) &= P(n|n-1)(P(n|n-1) + \sigma_v^2)^{-1}, \\ \hat{x}(n|n) &= \hat{x}(n|n-1) + K(n)(y(n) - \hat{x}(n|n-1)), \\ P(n|n) &= (1 - K(n))P(n|n-1). \end{aligned}$$

This works better than LMS or RLS, since we make use of the knowledge of the dynamics of $x(n)$.