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

EE 4C03 STATISTICAL DIGITAL SIGNAL PROCESSING AND MODELING

10 November 2016, 13:30–16:30

Open book exam: copies of the book by Hayes and the course slides allowed. No other materials allowed.

This exam has four questions (40 points).

Question 1 (9 points)

Consider the complex random process

$$x(n) = Ae^{j(\omega_0 n+\phi)} + w(n)$$

where $w(n)$ is a zero mean white Gaussian noise random process with variance σ_w^2. For each of the following cases,

– find the mean and the autocorrelation sequence of $x(n)$;

– if the process is wide sense stationary (WSS), find the power spectrum.

(a) A is a Gaussian random variable with zero mean and variance σ_A^2, and both ω_0 and ϕ are constants.

(b) ϕ is uniformly distributed over the interval $[-\pi, \pi]$ and both A and ω_0 are constants.

(c) ω_0 is a random variable that is uniformly distributed over some interval $[\Omega_0-\Delta, \Omega_0+\Delta]$, and both A and ϕ are constants.

(d) ω_0 is a random variable that is uniformly distributed over some interval $[\Omega_0-\Delta, \Omega_0+\Delta]$, ϕ is uniformly distributed over the interval $[-\pi, \pi]$, and A is a constant.

Hint: you may need this DTFT pair:

$$x(n) = \frac{\sin(\Delta n)}{\pi n} \quad \leftrightarrow \quad X(\omega) = \begin{cases} 1, & |\omega| < \Delta \\ 0, & \text{elsewhere} \end{cases}$$

Solution

(a) When ω_0 and ϕ are constants, then

$$\begin{aligned} m_x(n) &= E\{Ae^{j(\omega_0 n+\phi)} + w(n)\} \\ &= E\{A\}e^{j(\omega_0 n+\phi)} \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} r_x(k,l) &= E\{x(k)x^*(l)\} \\ &= E\{Ae^{j(\omega_0 k+\phi)}Ae^{-j(\omega_0 l+\phi)}\} + \sigma_w^2\delta(k-l) \\ &= E\{A^2\}e^{j\omega_0(k-l)} + \sigma_w^2\delta(k-l) \\ &= \sigma_A^2 e^{j\omega_0(k-l)} + \sigma_w^2\delta(k-l) \end{aligned}$$

This depends only on the difference $k-l$, and $x(n)$ is a WSS process. The power spectrum is

$$P_x(e^{j\omega}) = 2\pi\sigma_A^2 u_0(\omega-\omega_0) + \sigma_w^2$$

(b) Because ϕ varies uniformly over $[-\pi, \pi]$, again $m_x(n) = 0$, and

$$\begin{aligned} r_x(k,l) &= E\{A^2 e^{j(\omega_0(k-l)}\} + \sigma_w^2\delta(k-l) \\ &= A^2 e^{j\omega_0(k-l)} + \sigma_w^2\delta(k-l) \end{aligned}$$

The process is WSS, and

$$P_x(e^{j\omega}) = 2\pi A^2 u_0(\omega-\omega_0) + \sigma_w^2$$

(c) Now

$$m_x(n) = E\{Ae^{j(\omega_0 n+\phi)}\} = Ae^{j\phi}E\{e^{j\omega_0 n}\}$$

The expected value can be evaluated as

$$E\{e^{j\omega_0 n}\} = \frac{1}{2\Delta}\int_{\Omega_0-\Delta}^{\Omega_0+\Delta} e^{j\omega_0 n} d\omega_0 = e^{j\Omega_0 n}\frac{\sin(\Delta n)}{\Delta\, n}$$

The expected value is not independent of n, and $x(n)$ is not WSS.

The autocorrelation is

$$\begin{aligned} r_x(k,l) &= A^2 E\{e^{j\omega_0(k-l)}\} + \sigma_w^2\delta(k-l) \\ &= A^2 e^{j\Omega_0(k-l)}\tfrac{\sin(\Delta(k-l))}{\Delta\,(k-l)} + \sigma_w^2\delta(k-l) \end{aligned}$$

Since it is not WSS, the power spectrum is not defined.

(d) If now ϕ is uniformly varying, we obtain again $m_x(n) = 0$, while the autocorrelation is as before only dependent on the difference $k-l$, hence

$$r_x(k) = A^2 e^{j\Omega_0 k}\tfrac{\sin(\Delta k)}{\Delta\, k} + \sigma_w^2\delta(k)$$

The process is WSS. Using the given DTFT pair, the power spectrum is

$$P_x(\omega) = \begin{cases} \frac{A^2\pi}{\Delta} + \sigma_w^2, & \Omega_0-\Delta < \omega_0 < \Omega_0+\Delta \\ 0, & \text{elsewhere} \end{cases}$$

Question 2 (10 points)

Suppose that $x(n)$ is a wide-sense stationary process with autocorrelation sequence $r_x(n)$ and our goal is to model $x(n)$ using an optimal AR (all-pole) model with a system transfer function of the form

$$H(z) = \frac{b(0)}{1+\sum_{k=1}^{p} a_p(k)z^{-k}}\,,$$

where p is the model order. We consider the Levinson-Durbin recursion and the corresponding FIR lattice filter.

(a) Write down the Yule-Walker equations for this scenario.

(b) What is the computational complexity (order of magnitude) of the Levinson method as a function of the filter order p? And what is the complexity of solving the Yule-Walker equations directly?

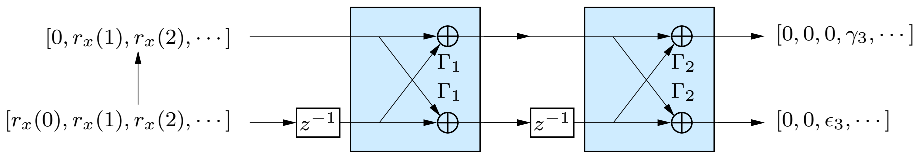
(c) Assume that the autocorrelation function of $x(n)$ is given by

$$r_x(-3) = 0,\ r_x(-2) = 1,\ r_x(-1) = 2,\ r_x(0) = 3,\ r_x(1) = 2,\ r_x(2) = 1,\ r_x(3) = 0.$$

Determine the reflection coefficients Γ_j, the model parameters $a_j(k)$ and the modeling errors ϵ_j for $j = 1, 2$.

(d) Conversely, given the reflection coefficients, it should be possible to recover the autocorrelation sequence.

Recall that the Schur recursion resulted in a FIR lattice filter that showed how the reflection coefficients are obtained from the autocorrelation sequence:



Starting from this, derive a lattice filter implementation that, given the reflection coefficients and $r_x(0)$, generates the rest of autocorrelation sequence at one of its outputs. (Specify also the input of the filter.)

Now suppose that $x(n)$ is a wide-sense stationary process and our goal is to predict $x(n)$ one step ahead. In other words, we want to estimate $x(n+1)$ from $x(n)$, $x(n-1)$, ..., $x(n-p)$ using a linear filter $w(n)$ of order p, i.e., $\hat{x}(n+1) = w(0)x(n)+w(1)x(n-1)+\cdots+w(p)x(n-p)$.

(e) Write down the Wiener-Hopf equations for this scenario (without solving them).

(f) How is the solution for the filter coefficients $w(n)$ related to the optimal all-pole model for the random process $x(n)$? Write down the system transfer function for this all-pole model using the filter coefficients $w(n)$.

Solution

(a)

$$\begin{bmatrix} r_x(0) & r_x(-1) & \cdots & r_x(-p) \\ r_x(1) & r_x(0) & \vdots & r_x(-p+1) \\ \vdots & \vdots & & \vdots \\ r_x(p) & r_x(p-1) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ \vdots \\ a_p(p) \end{bmatrix} = \begin{bmatrix} |b(0)|^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or (focusing only on the a-coefficients)

$$\begin{bmatrix} r_x(0) & r_x(-1) & \cdots & r_x(-p+1) \\ r_x(1) & r_x(0) & \vdots & r_x(-p+2) \\ \vdots & \vdots & & \vdots \\ r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} a_p(1) \\ \vdots \\ a_p(p) \end{bmatrix} = - \begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(p) \end{bmatrix}$$

(b) The computational complexity of the Levinson method is $\mathcal{O}(p^2)$, while directly solving the Yule-walker equations costs $\mathcal{O}(p^3)$.

(c) The solution can be obtained using the following steps:
Step 1:

$$a_0(0) = 1, \quad \epsilon_0 = r_x(0) = 3$$

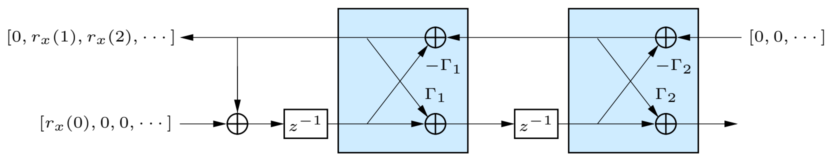
Step 2:

$$\begin{aligned}
&\gamma_0 = r_x(1) = 2 \\
&\Gamma_1 = -\gamma_0/\epsilon_0 = -2/3 \\
&a_1(1) = \Gamma_1 = -2/3 \\
&\epsilon_1 = \epsilon_0(1-\Gamma_1^2) = 3(1-4/9) = 5/3
\end{aligned}$$

Step 3:

$$\begin{aligned}
&\gamma_1 = r_x(2) + a_1(1) r_x(1) = 1 - 2/3 \cdot 2 = -1/3 \\
&\Gamma_2 = -\gamma_1/\epsilon_1 = 1/5 \\
&a_2(1) = a_1(1) + \Gamma_2 a_1(1) = -2/3 - 1/5 \cdot 2/3 = -4/5 \\
&a_2(2) = \Gamma_2 = 1/5 \\
&\epsilon_2 = \epsilon_1(1-\Gamma_2^2) = 5/3(1-1/25) = 8/5
\end{aligned}$$

(d) This is obtained by 'reversing' the direction of the top arrows. The copying of the autocorrelation coefficients is replaced by an addition. This gives



At the right side, the input is set to zero and the output is not used. This gives the 'maximum entropy' solution. Instead, the output can be fed back to that input via more lattice sections and arbitrary chosen reflection coefficients $|\Gamma_j| < 1$, $j = 3, 4, \cdots$, showing that the recovered autocorrelation sequence is not unique for $r_x(j)$, $j = 3, 4, \cdots$.

(e) The (real-valued) Wiener-Hopf equations are of the form:

$$\begin{bmatrix} r_x(0) & r_x(1) & \cdots & r_x(p) \\ r_x(1) & r_x(0) & \cdots & r_x(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_x(p) & r_x(p-1) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ \vdots \\ w(p) \end{bmatrix} = \begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(p+1) \end{bmatrix}.$$

(f) This resembles the Yule-Walker equations for all-pole modeling if we replace $a(n)$ by $-w(n-1)$. Hence, the all-pole system function related to the solution for $w(n)$ will be of order $p+1$ and it is given by

$$H(z) = \frac{b(0)}{1 - w(0)z^{-1} - w(1)z^{-2} \cdots - w(p)z^{-p-1}}.$$

Question 3 (12 points)

Consider a signal $s(n)$ that is obtained by filtering zero-mean white noise with variance 1. More specifically, assume that

$$s(n) = v(n) + v(n-1),$$

with $v(n)$ zero-mean white noise with variance 1. Also assume that this filtering process is corrupted by noise. In other words, we can only measure the signal

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

with $w(n)$ also zero-mean white noise with variance 1.

(a) Determine the unit sample response of the filter with input $v(n)$ and output $s(n)$. Also derive the frequency response of that filter and make a plot of the magnitude of the frequency response. What type of filter is this, low-pass, band-pass, or high-pass?

(b) Derive the autocorrelation function $r_s(k)$ of $s(n)$, the autocorrelation function $r_x(k)$ of $x(n)$, and the cross-correlation function $r_{sx}(k)$ between $s(n)$ and $x(n)$.

(c) Determine the unit sample response of a FIR filter of length 2 which optimally estimates the signal $s(n+1)$ from $x(n)$. Also compute the related minimal estimation error.

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

(d) Repeat (c) for a FIR filter of length 3. So again determine the unit sample response and compute the estimation error. Is the filter of length 3 better than the filter of length 2?

Hint: In a first step, try to use the last Wiener-Hopf equation to transform the system of three equations into a system of two equations. Then solve this as before.

(e) Write down the update equations of the normalized LMS (NLMS) algorithm to solve this estimation problem adaptively (you can either take the filter of length 2 or 3 to explain).

(f) Is there any difference between estimating $s(n+1)$ from $x(n)$ or estimating $x(n+1)$ from $x(n)$? Explain why there is a difference or why not.

Solution

(a) The unit sample response is given by

$$h(n) = \delta(n) + \delta(n-1).$$

The frequency response is given by

$$H(e^{j\omega}) = 1 + e^{-j\omega} = 1 + \cos(\omega) - j\sin(\omega).$$

The magnitude square can be expressed as

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

The latter is 4 at $\omega = 0$ and drops down to 0 towards $\omega = -\pi$ and $\omega = \pi$. Hence, it clearly has a low-pass shape.

(b) The autocorrelation function of $s(n)$ is given by

$$r_s(k) = \ldots, 0, 1, 2, 1, 0, \ldots$$

The autocorrelation function of $x(n)$ is given by

$$r_x(k) = \ldots, 0, 1, 3, 1, 0, \ldots$$

Finally, the cross-correlation function between $s(n)$ and $x(n)$ is the same as the auto-correlation function of $s(n)$ since $w(n)$ is independent of $s(n)$. Hence we have

$$r_{sx}(k) = r_s(k).$$

(c) The Wiener-Hopf equations are given by

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

If we plug in the results from (b), we obtain

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

leading to the solution

$$\begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = 1/8 \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3/8 \\ -1/8 \end{bmatrix}.$$

The related estimation error is given by

$$\begin{aligned} \xi_{\min} &= r_d(0) - \sum_l w(l) r_{dx}(l) = r_s(0) - \sum_l w(l) r_s(l+1) \\ &= r_s(0) - w(0) r_s(1) - w(1) r_s(2) = 2 - 3/8.1 = 1.625 \end{aligned}$$

(d) The Wiener-Hopf equations to solve now look like

$$\begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ w(2) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

You can use the last equation to establish the relation $w(2) = -w(1)/3$. This allows us to reduce the set of equations into

$$\begin{bmatrix} 3 & 1 \\ 1 & 8/3 \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

leading to the solution

$$\begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = 1/7 \begin{bmatrix} 3 & 1 \\ 1 & 8/3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 8/21 \\ -1/7 \end{bmatrix}.$$

As a result the filter coefficients are given by

$$\begin{bmatrix} w(0) \\ w(1) \\ w(2) \end{bmatrix} = \begin{bmatrix} 8/21 \\ -1/7 \\ 1/21 \end{bmatrix}.$$

The related estimation error is given by

$$\begin{aligned} \xi_{\min} &= r_d(0) - \sum_l w(l) r_{dx}(l) = r_s(0) - \sum_l w(l) r_s(l+1) \\ &= r_s(0) - w(0)r_s(1) - w(1)r_s(2) - w(2)r_s(3) = 2 - 8/21.1 = 1.619 \end{aligned}$$

Although the filter of length 3 does not provide a major improvement, there is some gain.

(e) These are the standard equations

(f) There is no difference which is immediately clear from the Wiener-Hopf equations. The intuition is that since $w(n)$ is white noise, it cannot be predicted. Hence, the only part in $x(n+1) = s(n+1) + w(n+1)$ that can be predicted is the part related to $s(n+1)$, whereas the part related to $w(n+1)$ can never be predicted.

Question 4 (9 points)

A continuous-time signal $x_a(t)$ is bandlimited to 5 kHz, i.e., $x_a(t)$ has a spectrum $X_a(f)$ that is zero for $|f| > 5$ kHz. Only 10 seconds of the signal has been recorded and is available for processing. We would like to estimate the power spectrum of $x_a(t)$ using the avalable data with a DFT algorithm, and it is required that the estimate has a resolution of at least 10 Hz. Suppose that we use Bartlett's method of periodogram averaging.

(a) What is the Nyquist rate for this signal?

If the data is sampled at the Nyquist rate, what is the mimimum section length L that you may use to get the desired resolution?

(b) Using this minimum section length, with 10 seconds of data, how many sections are available for averaging?

(c) What is the variance of the resulting Bartlett's spectrum estimate?

(d) What happens to the variance if we sample at twice the Nyquist rate (still collecting 10 seconds of data and aiming for a resolution of 10 Hz)?

Is there a benefit in sampling faster than Nyquist?

(e) Now we sample at Nyquist rate, but consider a Welch power spectrum estimate with a Bartlett window and 50% overlap. If the estimate has a resolution of 10 Hz, then what is the variance of the Welch estimate?

Is there a benefit in using the Welch estimate compared to the Bartlett estimate? Does the choice of the window play any role?

Solution

(a) The Nyquist rate is $f_s = 10$ kHz.

A resolution of $\Delta f = 10$ Hz (in analog frequency) implies that we want a resolution (in radians) of

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

Since the resolution of the periodogram using an L-point data record is

$$\text{Res}[\hat{P}_{per}(e^{j\omega})] = \Delta\omega = 0.89 \frac{2\pi}{L}$$

then for Bartlett's method we want to use a section length of

$$L \geq 0.89 \frac{2\pi}{\Delta\omega} = 890 \text{ samples}$$

(b) Sampling at 10 kHz, 10 seconds of data corresponds to $N = 10^5$ samples. Using $L = 890$, we have

$$K = \lfloor \frac{N}{L} \rfloor = 112$$

(c)

$$\text{var}[\hat{P}_B(e^{j\omega})] \approx \frac{1}{K} P_x^2(e^{j\omega})$$

(d) If the sampling rate is increased then $\Delta\omega$ decreases which, in turn, requires a longer section length for a given resolution. However, an increase in the sampling rate produces a corresponding increase in the total number of samples N within the same 10 second interval. Therefore, the number of sections K is the same, and the variance is unchanged.

(Sampling faster than Nyquist enables subsequent downsampling, which involves low-pass filtering (i.e. averaging) and this does result in reduced variance.)

(e) The window affects the resolution. For a Bartlett window, the resolution is

$$\text{Res}[\hat{P}_W(e^{j\omega})] = \Delta\omega = 1.28 \frac{2\pi}{L}$$

Thus, for the same resolution we need L to be 1.43 times as large compared to the periodogram: $L = 1280$.

The variance is independent of the window. Due to the 50% overlap, we have more sections than before, but not twice as many due to the larger L:

$$K = \lfloor 2\frac{N}{L} - 1 \rfloor = 155.$$

The variance is determined by the amount of averaging (K) and a penalty 9/8 due to the dependence of the overlapping sections, and given by

$$\text{var}[\hat{P}_W(e^{j\omega})] \approx \frac{9}{8K} P_x^2(e^{j\omega}) \approx 0.81 \text{var}[\hat{P}_B(e^{j\omega})]$$

At the same resolution, the variance of Welch with a Bartlett window is 81% of the variance of Bartlett's method (no windowing).