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

EE4C03 STATISTICAL DIGITAL SIGNAL PROCESSING AND MODELING

25 January 2018, 18:30-21:30

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 (10 points)

Given the random signal $x[n]$:

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

where A is a deterministic amplitude, ω_0 is a deterministic frequency, ϕ is a uniformly distributed random phase, and $w[n]$ is an i.i.d. zero mean Gaussian noise with variance $\sigma_w^2 = 1$.

(a) Determine the corresponding autocorrelation sequence $r_x(k)$.

(b) Find the power spectral density $P_x(\omega)$, and draw a plot of $P_x(\omega)$.

(c) How can you estimate the frequency ω_0 from $P_x(\omega)$? What are two issues that complicate or limit this in practice?

(d) Find the autocorrelation matrix $\mathbf{R}_x$ of size 5×5.

(e) What are the eigenvalues of this matrix? What information do they give us on the signal?

(f) How can the frequency ω_0 be estimated from $\mathbf{R}_x$?

(g) In practice, we only have an estimate $\hat{\mathbf{R}}_x$. What is a suitable estimate $\hat{\mathbf{R}}_x$? Is the resulting frequency estimate consistent?

Solution

(a)
$$r_x(k) = E(x(n)x^*(n-k)) = A^2 e^{j\omega_0 k} + \sigma_w^2 \delta(k)$$

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

where $u_0(\omega)$ is a (periodic with 2π) delta spike in the ω domain.

(c) Look for the peak in the spectrum. Limitations: with a finite number of samples, the spectrum gets convolved in the frequency domain with the DFT of the window function, which broadens the peak (reduced resolution) and gives sidelobes (sidelobe confusion). Also, with few samples, the signal may drown in the "finite sample noise" (i.e., the variance of the spectrum estimate is high).

(d)

$$\mathbf{R}_x = A^2 \begin{bmatrix} 1 & e^{-j\omega_0} & e^{-2j\omega_0} & e^{-3j\omega_0} & e^{-4j\omega_0} \\ e^{j\omega_0} & 1 & e^{-j\omega_0} & e^{-2j\omega_0} & e^{-3j\omega_0} \\ e^{j2\omega_0} & e^{-j\omega_0} & 1 & e^{-j\omega_0} & e^{-2j\omega_0} \\ e^{j3\omega_0} & e^{j2\omega_0} & e^{-j\omega_0} & 1 & e^{-j\omega_0} \end{bmatrix} + \sigma_w^2 \mathbf{I}_{5\times 5}$$

$$= A^2 \begin{bmatrix} 1 \\ e^{j\omega_0} \\ e^{2j\omega_0} \\ e^{3j\omega_0} \\ e^{4j\omega_0} \end{bmatrix} [1 \quad e^{-j\omega_0} \quad e^{-2j\omega_0} \quad e^{-3j\omega_0} \quad e^{-4j\omega_0}] + \sigma_w^2 \mathbf{I}_{5\times 5}$$

(e) Rank 1 plus $\sigma_w^2 \mathbf{I}$: the eigenvalues are $\lambda_1 = 5A^2 + \sigma_w^2$; $\lambda_{2,\cdots,5} = \sigma_w^2$

This shows the number of signals, the noise power, and then also the signal power.

(f) There are several options. If we work directly with $\mathbf{R}_x$, and know there is only 1 sinusoidal signal in noise, we can take e.g.,

$$\hat{\omega}_0 = \text{angle}\{\frac{r_{3,1}}{r_{2,1}}\}$$

or the phase of any other ratio of two subsequent entries in $\mathbf{R}_x$ that does not involve the main diagonal. Letting $\mathbf{u}_0$ be the first (dominant) eigenvector, then we can also take

$$\hat{\omega}_0 = \text{angle}\{\frac{u_0(2)}{u_0(1)}\}$$

or any other ratio of two subsequent entries. We could also average such ratios.

A better performance is expected by looking at the peaks of

$$J(\omega) = \mathbf{a}^H(\omega)\mathbf{u}_0\mathbf{u}_0^H\mathbf{a}(\omega)\,, \qquad \mathbf{a}(\omega) = \begin{bmatrix} 1 \\ e^{j\omega} \\ e^{2j\omega} \\ e^{3j\omega} \\ e^{4j\omega} \end{bmatrix}$$

or, similarly, using the MUSIC algorithm, which in this case reduces to

$$J(\omega) = \frac{1}{\mathbf{a}^H(\omega)(\mathbf{I} - \mathbf{u}_0\mathbf{u}_0^H)\mathbf{a}(\omega)}\,.$$

(These latter techniques can be easily extended to the case of multiple sinusoids in noise.)

(g)

$$\hat{\mathbf{R}}_x = \frac{1}{N-5}\sum_{n=1}^{N-4} \mathbf{x}(n)\mathbf{x}(n)^H\,, \qquad \mathbf{x}(n) = \begin{bmatrix} x(n) \\ x(n+1) \\ \cdots \\ x(n+4) \end{bmatrix}$$

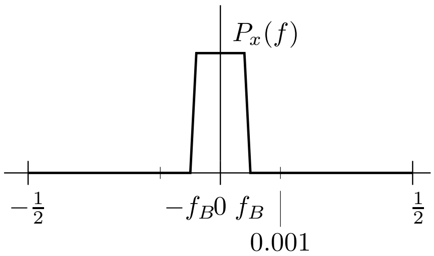
For this estimate, $E(\hat{\mathbf{R}}_x) = \mathbf{R}_x$ (unbiased), also as $N \to \infty$ we have $\hat{\mathbf{R}}_x \to \mathbf{R}_x$, hence the estimate is consistent.

Question 2 (8 points)

Suppose we wish to determine the bandwidth of a slowly varying stochastic process $x(n)$. An example is when we wish to determine the Doppler bandwidth f_B, and the speed, of a mobile terminal in a wireless communication system. In this case, the power spectrum $P_x(f)$ is nonzero only for frequencies up to f_B, i.e. it satisfies

$$P_x(f) = 0, \quad f_B < |f| \leq 1/2 \,.$$

We do not know f_B exactly, but we do know that $0 < f_B < 10^{-3}$.



We choose to compute an estimate $\hat{P}_x(f)$ of $P_x(f)$ from the data $x(n)$, plot $\hat{P}_x(f)$, and read the value of f_B from the plot.

(a) Suppose that $\hat{P}_x(f)$ is obtained by Bartlett's method with a window length of L. Suggest an appropriate value for L so that we can clearly read the value of f_B from our plot of $\hat{P}_x(f)$ (e.g., such that we reach a resolution of 10% of the maximum bandwidth). You may assume that the number of data samples is large, i.e., $N \gg L$.

(b) What is the variance of $\hat{P}_x(f)$ obtained using Bartlett's method?

(c) Suppose that we wish to achieve that the variance of $\hat{P}_x(f)$ is 10% of $P_x^2(f)$. How many samples of the signal are needed, and how is the spectrum to be estimated?

(d) If we wish to reduce the value of L, and the length of the FFTs used in Bartlett's method, we could downsample $x(n)$ before estimating the spectrum. The downsampled signal is $y(n) = x(nD)$, with D a positive integer.

The spectrum of the downsampled signal $y(n) = x(nD)$ satisfies

$$P_y(f) = \frac{1}{D} P_x(f/D) \tag{1}$$

as long as aliasing does not occur.

How large can we choose the downsampling factor D without losing the possibility to estimate f_B from the spectrum of $y(n)$?

What is after downsampling a suitable choice for L to reach again a resolution of 10% of the maximum bandwidth?

Solution

(a) Theory tells us that $E\{\hat{P}_x(f)\} = P_x(f) * W_B(f)$. Thus, if the 3dB width of the main lobe of $W_B(f)$, which is (using $\omega = 2\pi f$ to connect to Hayes),

$$\Delta\omega = \frac{0.89}{L}$$

is small in relation to the bandwidth of $P_x(f)$, changing f_B will not significantly alter the appearance of $P_x(f) * W_B(f)$. In other words, the resolution of the spectrum estimation must be sufficiently large in relation to what we try to see in the spectrum. Setting

$$\Delta\omega = \frac{0.89}{L} \leq 10^{-3} \cdot 0.10$$

yields that $L \geq 8900$ in order to get the required resolution.

(b) Let $K = N/L$, then

$$\text{var}(\hat{P}_B(f)) \approx \frac{1}{K} P_x^2(f)$$

(c) In this case, $K = 10$, hence $N = KL \geq 89000$.

The signal is split into K segments of L samples, we apply an FFT to each segment, take the amplitude squared, and average the K resulting segments.

$$\hat{P}_B(f) = \frac{1}{N} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} x(n + iL) e^{-j2\pi fn} \right|^2$$

(In practice we would also apply a window to improve the spectral characteristics: Welch method.)

(d) Aliasing does not occur if $f \geq \frac{1}{2D}$. Thus, if we should be able to represent all frequencies $f \leq 10^{-3}$, it should hold that $D \leq 500$, i.e., we can downsample the signal by up to a factor 500. Doing so would save much computation in the spectrum estimation. We obtain

$$\frac{0.89}{L} \leq 0.5 \cdot 0.10$$

hence $L \geq 17.8$. We would probably choose $L = 32$ (to be a power of 2).

Question 3 (12 points)

One problem in spectrum estimation concerns the issue whether or not a finite length sequence,

$$r_x(0), r_x(1), r_x(2), r_x(3), \ldots, r_x(p)$$

can be extended or extrapolated into a legitimate autocorrelation sequence in such a way that

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k) e^{-jk\omega} \tag{2}$$

is a valid power spectrum. This is also known as the extendibility problem. To make it more concrete, is it possible to find values of $r_x(k)$ for $|k| > p$ such that $P_x(e^{j\omega})$ in (2) is a non-negative real function of ω?

(a) When is a $(p+1) \times (p+1)$ matrix $\mathbf{R}_x$ a valid autocorrelation matrix? How can you describe the extendibility problem in terms of conditions on $\mathbf{R}_x$?

(b) Alternatively, what is the condition on the p reflection coefficients, obtained from the Levinson recursion, which guarantees that $r_x(0), r_x(1), r_x(2), \ldots, r_x(p)$ are $p+1$ valid samples from an autocorrelation sequence?

(c) How can you extend the list of reflection coefficients in such a way that the related autocorrelation sequence is a legitimate autocorrelation sequence? Is the extension unique?

(d) Use the procedure developed in (b) and (c) to determine constraints on a and b that are required for the sequence

$$r_x(0) = 1, \quad r_x(1) = a, \quad r_x(2) = b$$

to be an extendible autcorrelation sequence.

(e) Give a possible extension for $a = 0.3$ and $b = 0.9$, and explain how you obtain this extension.

Extra (this was not on the exam): Same for $a = 0.3$ and $b = 0.09$.

Solution

(a) The covariance matrix should be a symmetric positive definite matrix. So we need to extend the autocorrelation sequence such that this property of the covariance matrix is maintained. (It is not straightforward to do this directly on the covariance matrix, hence item (b)–(c).)

(b) The reflection coefficients should be smaller than 1 in magnitude, i.e., $|\Gamma_j| < 1$, for $j = 1, 2, \ldots, p$.

(c) One possibility that always works and does not introduce "new information" is to extend the sequence of reflection coefficients with zeros. So we can construct the sequence

$$\Gamma_j^e = \begin{cases} \Gamma_j, & j = 1, 2, \ldots, p \\ 0, & j > p \end{cases},$$

which corresponds to an all-pole power spectrum. However, the extension is not unique. It is possible to extend with any sequence such that $|\Gamma_j^e| < 1$.

(d) We require

$$|\Gamma_1| = |a| < 1$$

and

$$|\Gamma_2| = \left| \frac{r_x(2) + a_1(1) r_x(1)}{\epsilon_1} \right| = \left| \frac{r_x(2) + \Gamma_1 r_x(1)}{r_x(0)(1 - \Gamma_1^2)} \right| = \left| \frac{b - a^2}{1 - a^2} \right| < 1$$

or

$$2a^2 - 1 < b < 1.$$

(e) If $a = 0.3$ and $b = 0.9$, then $\Gamma_1 = -0.3$ and $\Gamma_2 = -0.89$. Setting $\Gamma_3 = 0$, we obtain the following related second-order filter: $a_3(0) = 1$, $a_2(1) = -0.033$, $a_3(2) = -0.89$, and $a_3(3) = 0$, and the error will be $\epsilon_3 = \epsilon_2 = 0.19$. This information can then be used to compute the $r_x(3)$ using the inverse Levinson-Durbin recursion.

If $a = 0.3$ and $b = 0.09$, then $\Gamma_2 = 0$, which means the autocorrelation sequence (from 0 up to 2) corresponds to the following all-pole filter of order 1:

$$H(z) = \frac{1}{1 - 0.3z^{-1}},$$

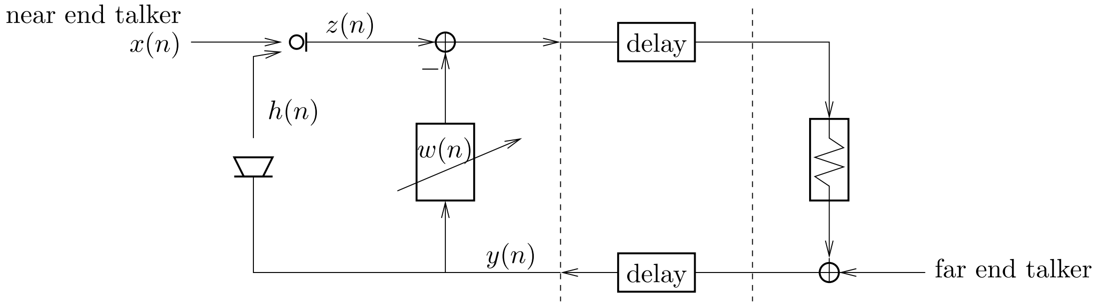
which generates the autocorrelation sequence $r_x(n) = 0.3^{|n|}$. If we further take $\Gamma_j = 0$ for $j > 2$, then this filter does not change and the extensions for the autocorrelation function are given by 0.3^3, 0.3^4, ...

Question 4 (10 points)

Consider the following echo cancellation problem for a hands free telephone system. The signal of interest that is picked up by the microphone is $x(n)$ (the near end signal) whereas the signal that is coming out of the speakers from the other end is $y(n)$ (the far end signal). Hence, the overall signal that is picked up by the microphone is given by

$$z(n) = x(n) + h(n) \star y(n) = x(n) + \sum_{l=0}^{L} h(l)y(n-l),$$

where $h(n)$ represents the propagation channel from the speaker to the microphone. Sending $z(n)$ to the other end is not a good idea since it will cause an echo. Hence, the problem is to find an estimate of $x(n)$ given only the signals $y(n)$ and $z(n)$, which are the only two signals known to the hands free telephone system, and which are assumed to be uncorrelated.



We will try to solve this echo cancellation problem by optimally estimating $z(n)$ from $y(n)$ using a linear filter $w(n)$.

(a) Derive the Wiener-Hopf equations for a linear filter $w(n)$ of order p such that if we apply this filter $w(n)$ to $y(n)$, we obtain the best possible estimate for $z(n)$ in the mean-square error sense.

Note that you have to give a derivation here and not simply copy equations from the book.

(b) Prove that when the filter order p is larger or equal to L, i.e., $p \geq L$, the optimal filter $w(n)$ is given by $h(n)$ padded with $p - L$ zeros at the end. In that case, also give an expression for the minimum mean-square error between $z(n)$ and its estimate.

Hint: Use the fact that $x(n)$ and $y(n)$ are uncorrelated.

(c) How can you use this optimal filter to estimate the near end signal $x(n)$? Can $x(n)$ be recovered exactly; or are there conditions under which this is possible?

(d) What is an LMS-type adaptive algorithm to estimate the filter coefficients $w(n)$?

Solution

(a) Let us define the vectors $\mathbf{y}_n = [y(n), y(n-1), \ldots, y(n-p)]^T$ and $\mathbf{w} = [w(0), w(1), \ldots, w(p)]^T$. The optimal filter $\mathbf{w}$ is then obtained by minimizing $\xi = E\{|z(n) - \mathbf{w}^T\mathbf{y}_n|^2\}$, which can be expressed as

$$\xi = \mathbf{w}^T E\{\mathbf{y}_n\mathbf{y}_n^T\}\mathbf{w} - 2\mathbf{w}^T E\{z(n)\mathbf{y}_n\} + E\{|z(n)|^2\}.$$

To minimize this quadratic expression, we can take the derivative towards $\mathbf{w}$ and set it to zero, which leads to

$$2E\{\mathbf{y}_n\mathbf{y}_n^T\}\mathbf{w} - 2E\{z(n)\mathbf{y}_n\} = \mathbf{0}.$$

Hence, the optimal filter $\mathbf{w}$ is the solution to the following Wiener-Hopf equations:

$$\mathbf{R}_y \mathbf{w} = \mathbf{r}_{zy},$$

where $\mathbf{R}_y = E\{\mathbf{y}_n \mathbf{y}_n^T\}$ and $\mathbf{r}_{zy} = E\{z(n)\mathbf{y}_n\}$.

(b) Let us first introduce the notation $\mathbf{h} = [h(0), h(1), \ldots, h(L)]^T$ and $\tilde{\mathbf{y}}_n = [y(n), y(n-1), \ldots, y(n-L)]^T$, such that $z(n) = x(n) + \mathbf{h}^T \tilde{\mathbf{y}}_n$. Then, since $x(n)$ and $y(n)$ are uncorrelated, we have

$$\mathbf{r}_{zy} = E\{z(n)\mathbf{y}_n\} = E\{\mathbf{y}_n \tilde{\mathbf{y}}_n^T\}\mathbf{h} = \tilde{\mathbf{R}}_y \mathbf{h},$$

where $\tilde{\mathbf{R}}_y = E\{\mathbf{y}_n \tilde{\mathbf{y}}_n^T\}$. Note that if $p = L$, then $\tilde{\mathbf{y}}_n = \mathbf{y}_n$ and $\tilde{\mathbf{R}}_y = \mathbf{R}_y$. As a result, the Wiener-Hopf equations turn into

$$\mathbf{R}_y \mathbf{w} = \tilde{\mathbf{R}}_y \mathbf{h}.$$

From this equation it is clear that as soon as $p \geq L$, the optimal solution for $\mathbf{w}$ is given by $\mathbf{h}$ padded with $p - L$ zeros at the bottom. If $p = L$, we obtain the optimal filter $\mathbf{w} = \mathbf{h}$.

If $p \geq L$, the output of the optimal filter is given by $\hat{z}(n) = \mathbf{w}^T \mathbf{y}_n = \mathbf{h}^T \tilde{\mathbf{y}}_n$, and thus the error is $e(n) = z(n) - \hat{z}(n) = x(n)$. Thus the minimum mean-square error is given by

$$\xi_{\min} = E\{|x(n)|^2\}.$$

(c) From the answer to (b), we see that if $p \geq L$, we have $e(n) = x(n)$. Hence, we use the error signal as our estimate of $x(n)$.

(d) The LMS-type adaptive filter is given by

$$\begin{aligned}
\mathbf{w}_{n+1} &= \mathbf{w}_n - \mu[(\mathbf{y}_n \mathbf{y}_n^T)\mathbf{w}_n - z(n)\mathbf{y}_n] \\
&= \mathbf{w}_n - \mu \mathbf{y}_n (\mathbf{w}_n^T \mathbf{y}_n - z(n)) \\
&= \mathbf{w}_n + \mu \mathbf{y}_n e(n),
\end{aligned}$$

if we define the error $e(n)$ as in the figure, i.e., $e(n) = z(n) - \mathbf{w}_n^T \mathbf{y}_n$.