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

EE4C03 STATISTICAL DIGITAL SIGNAL PROCESSING AND MODELING

24 January 2019, 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 (8 points)

Given the power spectral density:

$$P_x(e^{j\omega}) = 5 + 4\cos(\omega)$$

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

(b) Find a linear filter whose output process has this autocorrelation sequence, when excited by white noise.

(c) Is this filter a minimum-phase filter? Is this necessarily so?

(d) What is the variance of the output process?

Solution

(a) Expanding $P_x(e^{j\omega})$ in terms of complex exponentials,

$$P_x(e^{j\omega}) = 5 + 4\cos(\omega) = 5 + 2e^{j\omega} + 2e^{-j\omega}$$

it follows that $r_x(0) = 5$ and $r_x(1) = r_x(-1) = 2$. Thus

$$r_x(k) = 5\delta(k) + 2\delta(k-1) + 2\delta(k+1)$$

(b) Spectral factorization:

$$P_x(z) = \sigma_0^2 H(z) H^*(1/z^*)$$

Here:

$$P_x(z) = 5 + 2z^{-1} + 2z = (1 + 2z^{-1})(1 + 2z)$$

Thus a causal filter is $H(z) = 1 + 2z^{-1}$. (Alternative: $H(z) = 1 + \frac{1}{2}z^{-1}$.)

(c) Location of the zero: $z = -2$, outside the unit circle: not minimum phase.

It doesn't have to be: we could also choose $H(z) = 1 + \frac{1}{2}z^{-1}$ and $\sigma_0^2 = 4$ and obtain the same spectrum:

$$P_x(z) = 4(1 + \frac{1}{2}z^{-1})(1 + \frac{1}{2}z) = 5 + 2z^{-1} + 2z$$

This $H(z)$ is minimum phase (zero at $z = -\frac{1}{2}$). We would usually choose $H(z)$ to be minimum phase because then it can be stably inverted.

(d) $\sigma_x^2 = r_x(0) = 5$.

Question 2 (10 points)

Let $x(n)$ be a zero-mean real-valued wide-sense stationary random process with autocorrelation sequence $r_x(k) = 0.5^{|k|}$. This process is modeled by an all-pole filter, excited by white noise.

(a) Write down the expression for an all-pole filter of order 2 for $x(n)$, and the related Yule-Walker equations for finding the coefficients of the model.

(b) Solve the Yule-Walker equations using the Levinson recursion. For every iteration step, specify the corresponding modeling error $E\{(x(n) - \hat{x}(n))^2\}$ and reflection coefficients. Give the final model parameters and modeling error for the all-pole model of order 2.

(c) Is the resulting model stable? Why?

(d) What is the smallest order of an all-pole model for $x(n)$, above which the modeling error does not improve? Is this minimum modeling error equal to zero? If so, why? And if not, why is it not possible to reduce the modeling error to zero by increasing the model order?

Solution

(a) The all-pole model of order 2 is given by the filter $h(n)$ with z-transform

$$H(z) = \frac{b_0}{1 + a_1 z^{-1} + a_2 z^{-2}}.$$

The related Yule-Walker equations are given by

$$r_x(k) + a_1 r_x(k-1) + a_2 r_x(k-2) = |b_0|^2 \delta(k), \ \ k \geq 0.$$

The coefficients a_1 and a_2 are completely determined by the Yule-Walker equations for $k = 1, 2$, which can be summarized in matrix form as

$$\begin{bmatrix} r_x(0) & r_x(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = - \begin{bmatrix} r_x(1) \\ r_x(2) \end{bmatrix} \quad \Leftrightarrow \quad \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = - \begin{bmatrix} 0.5 \\ 0.25 \end{bmatrix}$$

(From the equation, it can already be seen that the solution is $a_1 = -0.5$, $a_2 = 0$.)

Since the process is real-valued, b_0 should be real and its value is determined by the Yule-Walker equation at $k = 0$:

$$b_0 = \sqrt{r_x(0) + a_1 r_x(1) + a_2 r_x(2)}.$$

(b) For order 0, we have $\epsilon_0 = r_x(0) = 1$.

For order 1, we have $\gamma_0 = r_x(1) = 0.5$, $\Gamma_1 = -\gamma_0/\epsilon_0 = -0.5$, $a_1(1) = \Gamma_1 = -0.5$, and $\epsilon_1 = \epsilon_0 + \Gamma_1 \gamma_0 = 0.75$.

For order 2, we have $\gamma_1 = r_x(2) + a_1(1) r_x(1) = 0$, $\Gamma_2 = -\gamma_1/\epsilon_1 = 0$, $a_2(1) = a_1(1) + \Gamma_2 a_1(1) = -0.5$, $a_2(2) = \Gamma_2 = 0$, and $\epsilon_2 = \epsilon_1 + \Gamma_2 \gamma_1 = 0.75$.

Also, $b(0) = \sqrt(0.75) = 0.866$. The resulting model is

$$H(z) = \frac{0.866}{1 - 0.5 z^{-1}}$$

Note that the model of order 2 is actually a model of order 1, since $a_2(2) = 0$ and $\epsilon_2 = \epsilon_1$, which makes sense since it is well-known that an autocorrelation sequence of the form $r_x(k) = \alpha^{|k|}$ corresponds to a first-order all-pole model for $x(n)$.

(c) Yes, the model is stable. This is true because the reflection coefficients Γ_1 and Γ_2 are smaller than 1 (Schur-Cohn stability test), or simply because the only pole of the filter equals $\alpha = 0.5$ which is within the unit circle.

(d) As already indicated, the optimal order is equal to 1. The minimum modeling error is not zero. The Yule-Walker equations basically realize a matching of the correlation functions, and that can be obtained perfectly with a first-order all-pole model. However, the modeling error ϵ actually corresponds to the prediction error between $x(n)$ and its prediction through $\hat{x}(n) = 0.5x(n-1)$ (which is the best possible linear prediction), and considering the random nature of $x(n)$ this prediction error is generally not zero. A zero prediction error can only be obtained if $x(n)$ is deterministic.

Question 3 (11 points)

Assume that we measure a real-valued signal $x(n)$ as $\bar{x}(n)$, and assume that the measurement process is perfect except for some *outliers*, i.e., $\bar{x}(n)$ is equal to $x(n)$ except for a few isolated values of n where $\bar{x}(n)$ is completely different from $x(n)$. Instead of eliminating those values, suppose that we perform a minimum mean-square interpolation as follows. If an outlier appears at time instant $n = n_0$, i.e., $|\bar{x}(n_0) - x(n_0)|$ is very large, we estimate $x(n_0)$ as

$$\hat{x}(n_0) = a\bar{x}(n_0 - 1) + b\bar{x}(n_0 + 1) = ax(n_0 - 1) + bx(n_0 + 1),$$

where the last equality is due to the fact that every outlier is isolated and thus we have no outliers at time instants $n_0 - 1$ and $n_0 + 1$.

(a) Assume that $x(n)$ is a zero-mean real-valued wide-sense stationary random process with autocorrelation sequence $r_x(k)$. Derive the Wiener-Hopf equations for the values of a and b that minimize the mean-square error

$$\xi = E\{|x(n_0) - \hat{x}(n_0)|^2\}.$$

Also derive an expression for the minimum mean-square error. Note that you have to give a *derivation* here and not simply copy equations from the book.

Hint: you will need the following matrix identity:

$$\begin{bmatrix} p & q \\ q & p \end{bmatrix}^{-1} = \frac{1}{p^2 - q^2} \begin{bmatrix} p & -q \\ -q & p \end{bmatrix}$$

(b) If $r_x(k) = 0.5^{|k|}$, evaluate the expressions for the optimal a and b, as well as the for the minimum mean-square error.

(c) Discuss when it may be better to use an estimator of the form

$$\hat{x}(n_0) = a\bar{x}(n_0 - 1) + b\bar{x}(n_0 - 2) = ax(n_0 - 1) + bx(n_0 - 2),$$

or explain why such an estimator will never be better.

(d) Suppose you want to develop a method to compute a and b adaptively. How would you design such a method? What would you use as input sequence for the adaptive method?

Solution

(a) Rewriting the estimate as $\hat{x}(n_0) = \mathbf{w}^T\mathbf{x}$, where $\mathbf{w} = [a, b]^T$ and $\mathbf{x} = [x(n_0-1), x(n_0+1)]^T$, we can express the mean-square error as

$$\begin{aligned}\xi &= E\{|x(n_0) - \mathbf{w}^T\mathbf{x}|^2\} \\ &= \mathbf{w}^T E\{\mathbf{x}\mathbf{x}^T\}\mathbf{w} - 2\mathbf{w}^T E\{\mathbf{x}x(n_0)\} + E\{|x(n_0)|^2\}. \end{aligned} \tag{1}$$

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

$$2\{\mathbf{x}\mathbf{x}^T\}\mathbf{w} - 2E\{\mathbf{x}x(n_0)\} = 0.$$

Hence, the minimum mean-square error filter $\mathbf{w}$ satisfies

$$E\{\mathbf{x}\mathbf{x}^T\}\mathbf{w} = E\{\mathbf{x}x(n_0)\}, \tag{2}$$

where we can express $E\{\mathbf{x}\mathbf{x}^T\}$ and $E\{\mathbf{x}x(n_0)\}$ in terms of $r_x(k)$ as

$$\begin{aligned}E\{\mathbf{x}\mathbf{x}^T\} &= \begin{bmatrix} r_x(0) & r_x(2) \\ r_x(2) & r_x(0) \end{bmatrix}, \\ E\{\mathbf{x}x(n_0)\} &= \begin{bmatrix} r_x(1) \\ r_x(1) \end{bmatrix}.\end{aligned}$$

As a result, we finally obtain

$$\begin{aligned}\mathbf{w} &= \begin{bmatrix} r_x(0) & r_x(2) \\ r_x(2) & r_x(0) \end{bmatrix}^{-1} \begin{bmatrix} r_x(1) \\ r_x(1) \end{bmatrix} \\ &= \frac{1}{r_x^2(0) - r_x^2(2)} \begin{bmatrix} r_x(0) & -r_x(2) \\ -r_x(2) & r_x(0) \end{bmatrix} \begin{bmatrix} r_x(1) \\ r_x(1) \end{bmatrix} \\ &= \frac{1}{r_x^2(0) - r_x^2(2)} \begin{bmatrix} r_x(1)[r_x(0) - r_x(2)] \\ r_x(1)[r_x(0) - r_x(2)] \end{bmatrix} \\ &= \frac{1}{r_x(0) + r_x(2)} \begin{bmatrix} r_x(1) \\ r_x(1) \end{bmatrix}.\end{aligned}$$

Plugging (2) into (1), we obtain the following expression for the minimum mean-square error:

$$\begin{aligned}\xi_{\min} &= E\{|x(n_0)|^2\} - \mathbf{w}^T E\{\mathbf{x}x(n_0)\} \\ &= E\{|x(n_0)|^2\} - E\{\mathbf{x}x(n_0)\}E\{\mathbf{x}\mathbf{x}^T\}^{-1}E\{\mathbf{x}x(n_0)\},\end{aligned}$$

which in terms of $r_x(n)$ can be written as

$$\xi_{\min} = r_x(0) - \frac{2r_x(1)^2}{r_x(0) + r_x(2)}.$$

(b) From the earlier expressions, it is easy to see that $a = b = 0.5/1.25 = 0.4$ and $\xi_{\min} = 0.6$.

(c) There are cases when it is better to use this estimator. Intuitively, it is clear that when $|r_x(2)| > |r_x(1)|$, we can obtain more information about $x(n_0)$ from $x(n_0 - 2)$ than from $x(n_0 + 1)$. As a result, in that case, the estimator $\hat{x}(n_0) = ax(n_0 - 1) + bx(n_0 - 2)$ will be better than the estimator $\hat{x}(n_0) = ax(n_0 - 1) + bx(n_0 + 1)$. This can also be shown mathematically.

(d) It is important that the outliers do not mess up the adaptive computations of a and b. Hence, in a first step we have to remove the outliers from $\bar{x}(n)$. More specifically, we have to look for all the time instants n_k for which $\bar{x}(n_k - 1)$, $\bar{x}(n_k)$, and $\bar{x}(n_k + 1)$ are free of outliers, and we then introduce the notation $x_k = x(n_k)$ and $\mathbf{x}_k = [x(n_k - 1), x(n_k + 1)]^T$. As before, we also introduce the notation $\mathbf{w}_k = [a_k, b_k]^T$, representing the value of the coefficients at time instant n_k. The updating process will now proceed as a function of k (assuming $n_k < n_{k+1}$). So we gradually increase k and compute new values for the coefficients. As soon as an outlier is detected (large residual error), we use the most recent values of a and b to carry out the interpolation.

In more detail, the LMS updating formula can then be written as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \mathbf{x}_k e_k = \mathbf{w}_k - \mu \mathbf{x}_k (\mathbf{w}_k^T \mathbf{x}_k - x_k).$$

A good initial condition for $\mathbf{w}$ is $\mathbf{w}_0 = \mathbf{0}$. The step size μ should be chosen such that we have convergence in the mean square:

$$0 < \mu < \frac{2}{E\{\|\mathbf{x}_k\|^2\}} = \frac{1}{r_x(0)}.$$

Alternatively, the updating formulas for the exponentially weighted RLS can be written as:

$$\begin{aligned}
\mathbf{P}_{k+1} &= \lambda^{-1}\mathbf{P}_k - \lambda^{-2}\frac{\mathbf{P}_k \mathbf{x}_{k+1}\mathbf{x}_{k+1}^T \mathbf{P}_k}{1 + \lambda^{-1}\mathbf{x}_{k+1}^T \mathbf{P}_k \mathbf{x}_{k+1}}, \\
\boldsymbol{\theta}_{k+1} &= \lambda \boldsymbol{\theta}_k + \mathbf{x}_{k+1} x_{k+1}, \\
\mathbf{w}_{k+1} &= \mathbf{P}_{k+1}\boldsymbol{\theta}_{k+1}.
\end{aligned}$$

Good initial conditions for $\boldsymbol{\theta}$ and $\mathbf{P}$ are for instance $\boldsymbol{\theta}_0 = \mathbf{0}$ and $\mathbf{P}_0 = 100\mathbf{I}$. The forgetting factor λ should be chosen within the range

$$0 < \lambda \leq 1.$$

A good choice is to take it very close to 1, e.g., $\lambda = 0.98$.

Question 4 (11 points)

Let $x(n)$ be a random process consisting of a single complex exponential in white noise, with covariance sequence

$$r_x(k) = P e^{jk\omega_0} + \sigma_w^2 \delta(k)$$

and $(p+1) \times (p+1)$ autocorrelation matrix $\mathbf{R}_x$.

In the context of minimum variance spectrum estimation, let $\mathbf{g}_i$ be a vector with the FIR filter coefficients of the minimum-variance bandpass filter,

$$\mathbf{g}_i = \frac{\mathbf{R}_x^{-1}\mathbf{e}_i}{\mathbf{e}_i^H \mathbf{R}_x^{-1}\mathbf{e}_i}, \qquad \mathbf{e}_i = [1, e^{j\omega_i}, e^{j2\omega_i}, \cdots, e^{jp\omega_i}]^T.$$

Denote by $G_i(e^{j\omega})$ the corresponding filter response.

(a) Explain how these bandpass filters are used in the estimation of the power spectrum of $x(n)$.

(b) Motivate that $G_i(e^{j\omega})$ is a bandpass filter that has center frequency ω_i with $G_i(e^{j\omega_i}) = 1$.

Hint: first determine and plot the filter response for a simple case where $\omega_i = 0$ and $\mathbf{R}_x = \mathbf{I}$. Then motivate what changes in more general cases.

(c) Using the matrix inversion lemma, we can derive that

$$\mathbf{R}_x^{-1} = \frac{1}{\sigma_w^2}\left(\mathbf{I} - \frac{\mathbf{e}_0\mathbf{e}_0^H P}{\sigma_w^2 + \mathbf{e}_0^H\mathbf{e}_0 P}\right).$$

Use this to determine a (very simple) expression for the filter response for $\omega_i = \omega_0$; give a plot of this response.

(d) Assuming that $\omega_i \neq \omega_0$, prove that $G_i(z)$ has a zero that approaches $z = e^{j\omega_0}$ as $\sigma_w^2 \to 0$.

(e) Give an explanation/interpretation for this.

Solution

(a) The filter is $G_i(\omega) = \mathbf{e}^H\mathbf{g}_i$, with $\mathbf{e} = [1, e^{j\omega}, e^{j2\omega}, \cdots, e^{jp\omega}]^T$. The filter relates to a 'filter bank' interpretation of spectrum analysis. The input signal $x(n)$ is applied to each filter $G_i(\omega)$, the output power $E(|y_i(n)|^2)$ of this filter (divided by the filter bandwidth) is the spectrum estimate for frequency ω_i.

We thus expect that $G_i(\omega)$ is a bandpass filter with a peak around $\omega = \omega_i$, and ideally no sidelobes. In practice, there are sidelobes, which means that energy from different frequencies plays a role in the power estimate at ω_i.

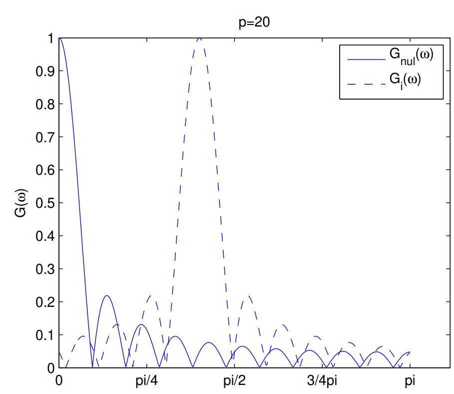
(b) For $\mathbf{R}_x = \mathbf{I}$, the FIR filter $G_{nul}(\omega)$ (for $\omega_i = 0$) has coefficients

$$\mathbf{g}_{nul} = \frac{\mathbf{1}}{\mathbf{1}^H\mathbf{1}} = \frac{1}{p+1}\mathbf{1}, \qquad \mathbf{1} := [1, 1, \cdots, 1]^T$$

which is a simple averaging filter with frequency response

$$G_{nul}(\omega) = \mathbf{e}^H\mathbf{g}_{nul} = \frac{1}{p+1}\sum_{k=0}^{p} 1 \cdot e^{-j\omega k} = \frac{1}{p+1}\frac{\sin(\omega(p+1)/2)}{\sin(\omega/2)}e^{j\omega(p+1)/2}$$

This is a Dirichlet kernel (periodic sinc function), centered around $\omega = 0$.



For other filters ω_i, the response is

$$G_i(\omega) = \frac{\mathbf{e}^H\mathbf{e}_i}{\mathbf{e}_i^H\mathbf{e}_i} = \frac{1}{p+1}\sum_{k=0}^{p} e^{j\omega_i k}e^{-j\omega k} = \frac{1}{p+1}\frac{\sin((\omega-\omega_i)(p+1)/2}{\sin((\omega-\omega_i)/2)}e^{j(\omega-\omega_i)(p+1)/2}$$

and hence simply shifts to center around ω_i. For more general $\mathbf{R}_x$, the responses deform; for this particular case in such a way that the frequency ω_0 is suppressed in case $\omega_i \neq \omega_0$.

(c)

$$G_0(e^{j\omega}) = \mathbf{e}^H\mathbf{g}_0 = \frac{\mathbf{e}^H\mathbf{R}_x^{-1}\mathbf{e}_0}{\mathbf{e}_0\mathbf{R}_x^{-1}\mathbf{e}_0} = \frac{\frac{\mathbf{e}^H\mathbf{e}_0}{\sigma_w^2} - \frac{\mathbf{e}^H\mathbf{e}_0(\mathbf{e}_0^H P\mathbf{e}_0)}{\sigma_w^2(\sigma_w^2+\mathbf{e}_0^H\mathbf{e}_0 P)}}{\cdots} = \mathbf{e}^H\mathbf{e}_0 \cdot [\text{normalization}] = \frac{\mathbf{e}^H\mathbf{e}_0}{p+1}$$

as the proper normalization is found e.g. for $\omega = \omega_0$, i.e., replacing $\mathbf{e}$ by $\mathbf{e}_0$.

This response is the Dirichlet function centered around ω_0.

(d) Similarly, the frequency response for the filter centered around $\omega_i \neq \omega_0$ is

$$G_i(e^{j\omega}) = \mathbf{e}^H\mathbf{g}_i = \frac{\lambda}{\sigma_w^2}\big[\mathbf{e}^H\mathbf{e}_i - \frac{P}{\sigma_w^2 + (p+1)P}\mathbf{e}^H\mathbf{e}_0\mathbf{e}_0^H\mathbf{e}_i\big]$$

where $\lambda = \frac{1}{\mathbf{e}_i^H\mathbf{R}_x^{-1}\mathbf{e}_i}$ is a normalization. Evaluate this at ω_0 by replacing $\mathbf{e}$ by $\mathbf{e}_0$:

$$\begin{aligned} G_i(e^{j\omega_0}) = \mathbf{e}_0^H\mathbf{g}_i &= \frac{\lambda}{\sigma_w^2}\big[\mathbf{e}_0^H\mathbf{e}_i - \frac{P}{\sigma_w^2 + (p+1)P}\mathbf{e}_0^H\mathbf{e}_0\mathbf{e}_0^H\mathbf{e}_i\big] \\ &= \frac{\lambda}{\sigma_w^2}\big[1 - \frac{(p+1)P}{\sigma_w^2 + (p+1)P}\big]\mathbf{e}_0^H\mathbf{e}_i \\ &= \frac{\lambda}{\sigma_w^2}\big[1 - \frac{(p+1)}{\frac{\sigma_w^2}{P} + (p+1)}\big]\mathbf{e}_0^H\mathbf{e}_i \end{aligned}$$

For $\frac{\sigma_w^2}{P} \to 0$, we find that $G_i(e^{j\omega_0}) \to 0$.

(e) This implies that all filters (except the one centered at ω_0) have a zero response at ω_0, so the energy of the sinusoid at ω_0 is suppressed when estimating the power at a different frequency. That is a special feature of the MV spectrum; a periodogram does not have that property. It will make the MV spectrum more 'spiky' around dominating frequencies (these do not cause sidelobes).