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

EE4C03 STATISTICAL DIGITAL SIGNAL PROCESSING AND MODELING

6 November 2019, 13:30–16:30

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

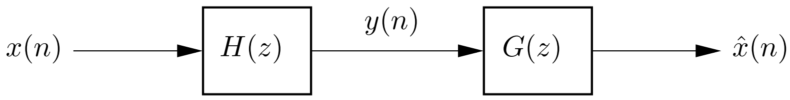
This exam has four questions (40 points)

Question 1 (10 points)

Consider a linear shift-invariant system with input $x(n)$ and output $y(n)$. The input-output relation is given by

$$y(n) = x(n) + \frac{1}{2}x(n-1) + \frac{1}{3}y(n-1).$$

The input $x(n)$ is real-valued zero-mean white noise with variance 1.



(a) Write down the expression for the system transfer function $H(z)$.

(b) What is the power spectral density $P_y(\omega)$ of $y(n)$?

(c) Give an expression for the autocorrelation sequence $r_y(k) = E\{y(n)y(n-k)\}$.

(d) Give an expression for the cross-correlation sequence between the input $x(n)$ and the output $y(n)$, i.e., $r_{xy}(k) = E\{x(n)y(n-k)\}$.

(e) Give expressions for the optimal FIR filter $G(z)$ of order 1 to estimate $x(n)$ from $y(n)$, given knowledge of the correlation functions. (Write this in sufficient detail; you do not have to evaluate the expressions numerically, but make clear how it could be done.)

(f) Does the filter under (e) provide an exact inverse of $H(z)$? How would you determine the best possible inverse filter?

(g) Does the optimal FIR filter $G(z)$ have to change if the input signal $x(n)$ is not white noise but some other random signal?

Solution

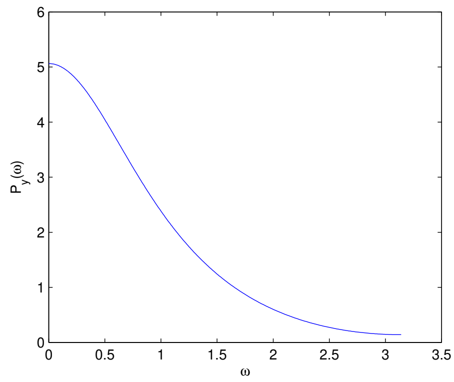
(a)

$$H(z) = \frac{1 + 1/2z^{-1}}{1 - 1/3z^{-1}}$$

(b)

$$P_y(z) = H(z)H^*(1/z^*) = \frac{1 + 1/2z^{-1}}{1 - 1/3z^{-1}} \frac{1 + 1/2z}{1 - 1/3z} = \frac{5/4 + 1/2z^{-1} + 1/2z}{10/9 - 1/3z^{-1} - 1/3z}$$

$$P_y(\omega) = \frac{5/4 + \cos(\omega)}{10/9 - 2/3\cos(\omega)}$$



(c) Follows from the inverse z-transform of $P_y(z)$. Use the transform pairs on Hayes p.17.

$$(\frac{1}{3})^{|k|} \quad \leftrightarrow \quad \frac{8/9}{(1-1/3z^{-1})(1-1/3z)}$$

Thus,

$$r_y(k) = \frac{5}{4}\frac{9}{8}\left(\frac{1}{3}\right)^{|k|} + \frac{1}{2}\frac{9}{8}\left(\frac{1}{3}\right)^{|k-1|} + \frac{1}{2}\frac{9}{8}\left(\frac{1}{3}\right)^{|k+1|}$$

This could be further simplified in various ways to the form $\alpha\delta(k) + \beta\left(\frac{1}{3}\right)^{|k|}$.

(d) The cross-correlation $r_{xy}(k)$ is given by

$$r_{xy}(k) = r_x(k) * h(-k)$$

A z-transform gives the relation

$$P_{xy}(z) = P_x(z)H(1/z) = \frac{1+1/2z}{1-1/3z}$$

Applying the inverse z-transform gives the result (anti-causal)

$$r_{xy}(k) = \left(\frac{1}{3}\right)^{-k} u(-k) + \frac{1}{2}\left(\frac{1}{3}\right)^{-k-1} u(-k-1)$$

Another option to obtain this result is to compute the impulse response of the filter from $H(z)$:

$$h(n) = (\frac{1}{3})^n u(n) + \frac{1}{2}(\frac{1}{3})^{n-1} u(n-1),$$

and to observe that $r_{xy}(k)$ is the time-reversed impulse response:

$$r_{xy}(k) = h(-k) = \left(\frac{1}{3}\right)^{-k} u(-k) + \frac{1}{2}\left(\frac{1}{3}\right)^{-k-1} u(-k-1).$$

(e) The optimal FIR filter $g(n)$ of order 1 can be computed using the Wiener-Hopf equations:

$$\begin{bmatrix} r_y(0) & r_y(-1) \\ r_y(1) & r_y(0) \end{bmatrix} \begin{bmatrix} g(0) \\ g(1) \end{bmatrix} = \begin{bmatrix} r_{xy}(0) \\ r_{xy}(1) \end{bmatrix}.$$

Using the expressions obtained in (c) and (d), this leads to

$$\begin{bmatrix} \frac{5}{4}\frac{9}{8} + \frac{3}{8} & \frac{5}{4}\frac{3}{8} + \frac{1}{2}\frac{1}{8} \\ \frac{5}{4}\frac{3}{8} + \frac{1}{2}\frac{1}{8} & \frac{5}{4}\frac{9}{8} + \frac{3}{8} \end{bmatrix} \begin{bmatrix} g(0) \\ g(1) \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ 0 \end{bmatrix},$$

which can be solved for $g(0)$ and $g(1)$ using a 2×2 matrix inverse.

(f) The filter does not provide an exact inverse, since the inverse of an IIR filter with poles and zeros is an IIR filter, and not an FIR filter. The best inverse filter is given by

$$G(z) = H^{-1}(z) = \frac{1 - 1/3z^{-1}}{1 + 1/2z^{-1}},$$

which is the filter given by

$$g(n) = (-\frac{1}{2})^n u(n) - \frac{1}{3}(-\frac{1}{2})^{n-1} u(n-1).$$

Note that since $h(n)$ is a causal stable minimum phase filter, the inverse is also a causal stable minimum phase filter.

Hayes ch.7.3 (IIR Wiener filter) describes how this filter is obtained from correlation data (we skipped that section in class).

(g) The optimal filter will not change, although the optimal first order FIR filter might change.

Question 2 (8 points)

Let $x(n)$ be a random process with autocorrelation sequence $r_x(k) = (0.2)^{|k|}$.

(a) What is the purpose of a prediction error filter?

(b) Find the reflection coefficients Γ_1 and Γ_2 for a second-order predictor related to $x(n)$.

(c) What conclusions can be drawn from the values of these reflection coefficients (e.g. if they are smaller/larger/equal to 1, or if they are zero).

(d) Draw a corresponding lattice filter implementation of the filter.

(e) Draw a filter implementation that, if the input is white noise $v(n)$, generates a signal with autocorrelation sequence $r_x(k)$ (also specify the values of the coefficients).

Solution

(a) A prediction error filter minimizes the prediction error of a linear prediction problem. It is an FIR filter $A_p(z)$ such that the output $E(z) = X(z)A_p(z)$ (the prediction error signal) is as close as possible to white noise: the prediction error sequence is uncorrelated up to lag p. Thus, $A_p(z)$ is used to "analyze" $X(z)$.

Conversely, the inverse filter $1/A_p(z)$ can be used to "synthesize" $X(z)$: by giving white noise as input to the inverse filter, a signal with the same correlations as $X(z)$ (up to lag p) will result. It can be interpreted as the denominator of an all-pole signal model. In other words, if $A_p(z)$ is the prediction error filter for $X(z)$, then $1/A_p(z)$ is a good model for $X(z)$ and filtering $X(z)$ with $A_p(z)$ results in some remaining noise that can be used to encode $X(z)$.

(b) See Hayes p.219 Table 5.1.

For order 0, we have $a_0(0) = 0$, $\epsilon_0 = r_x(0) = 1$.

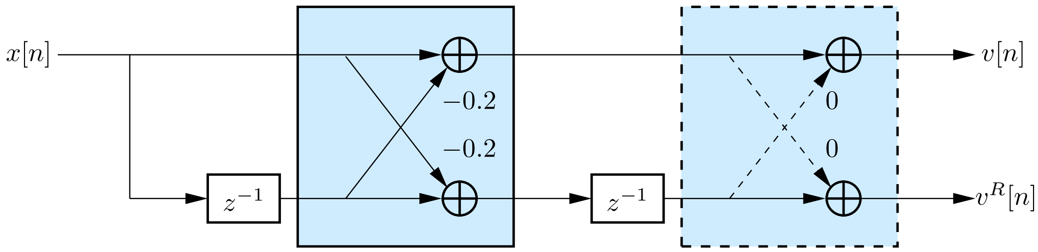
For order 1, we obtain $\gamma_0 = r_x(1) = 0.2$, $\Gamma_1 = -\frac{\gamma_0}{\epsilon_0} = -0.2$, $a_1(1) = -0.2$, and $\epsilon_1 = \epsilon_0 + \Gamma_1\gamma_0 = 0.96$.

For order 2, we obtain $\gamma_1 = r_x(2) + a_1(1)r_x(1) = 0$, $\Gamma_2 = -\frac{\gamma_1}{\epsilon_1} = 0$, $a_2(1) = -0.2$, $a_2(2) = 0$, and $\epsilon_2 = \epsilon_1 + \Gamma_2\gamma_1 = 0.96$.

(c) The fact that $\Gamma_2 = 0$ tells us that the optimal model was already obtained at order 1 and increasing the order does not improve the modeling (or reduce the error of the prediction error filter); we can stop the recursion. The fact that $|\Gamma_1| < 1$ tells us that the filter $1/A_p(z)$ is stable. If $|\Gamma_1| > 1$ the filter $1/A_p(z)$ would be unstable.

If the last reflection coefficient would be equal to 1, then the process $x(n)$ is deterministic (e.g., a constant or a sinusoid) and exactly predictable from p past samples without residual error. In this case, the recursion stops as well.

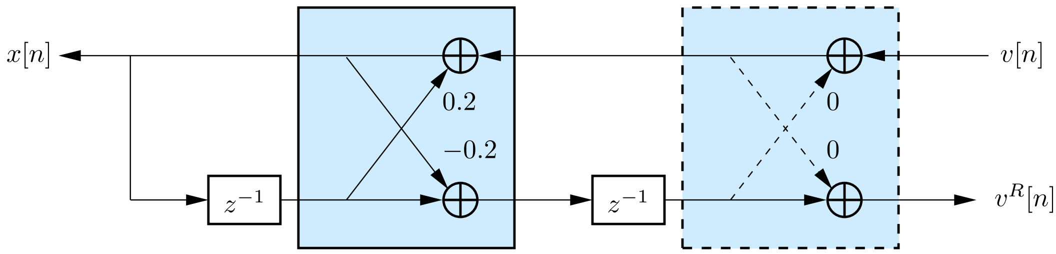
(d)



The prediction error is called $v[n]$ here (called $e[n]$ before).

Note from the figure it is easy to verify that $V(z) = X(z)(1 - 0.2z^{-1})$. We know from the given $r_x(k)$ that the spectrum of $X(z)$ is $P_x(z) = \frac{scale}{(1-0.2z^{-1})(1-0.2z)}$. Hence the spectrum of $V(z)$ is $P_v(z) = X(z)A(z)A^*(1/z^*) = \frac{scale}{(1-0.2z^{-1})(1-0.2z)} \cdot (1 - 0.2z^{-1}) \cdot (1 - 0.2z) = (scale)$, so we verified that we indeed obtain white noise. The scale is $1 - (0.2)^2$.

(e)



This is obtained by "reversing" the signal flow of $v[n]$ to $x[n]$ in the previous figure while maintaining the same relations between variables.

From the figure it is easy to verify that $X(z) = \frac{V(z)}{1-0.2z^{-1}}$, with correlation sequence $r_x(k) = \beta(0.2)^{|k|}$ for some scaling β that also depends on the power of the white noise.

Question 3 (10 points)

Consider a real-valued signal $x(n)$, which can be modeled as

$$x(n) = c\, d(n) + v(n),$$

where c is a real-valued deterministic parameter and $d(n)$ and $v(n)$ are real-valued mutually uncorrelated random processes with zero mean and respective variances σ_d^2 and σ_v^2.

We attempt to reconstruct $d(n)$ from $x(n)$ using the simple (zero-order) filtering operation

$$\hat{d}(n) = w\, x(n)$$

where we will adapt the coefficient w until it is optimal.

(a) Under which circumstances is a zero-order filter appropriate?

(b) Give the expression of the optimal Wiener filter w_{opt} and explain what happens if σ_v^2 is very small or very large.

(c) Give the update equation for the RLS filter, and try to write the new estimate of the filter $w(n+1)$ as a function of the previous estimate $w(n)$. Clearly define all the notations that you use.

(d) Assume now the following state space model

$$\begin{aligned} w(n+1) &= w(n), \\ d(n+1) &= x(n+1)w(n+1) + e(n+1), \end{aligned}$$

where we assume that $e(n)$ is a random process with zero mean and variance $\sigma_e^2 = 1$. Derive the Kalman filter for this state space model and again clearly define all used notation. Show that this filter is equivalent to the RLS filter derived earlier.

(e) How should the above state space model be adapted to make it equivalent to the exponentially weighted RLS filter?

Solution

(a) To invert another zero-order filter.

(b) The optimal Wiener filter is given by

$$w_{\text{opt}} = \frac{E\{x(n)d(n)\}}{E\{x^2(n)\}} = \frac{c\sigma_d^2}{c^2\sigma_d^2 + \sigma_v^2}.$$

If the noise is very small, the filter simply is $w_{\text{opt}} = 1/c$ which is indeed the correct operation for the noiseless model $x(n) = cd(n)$. If the noise is very large, the only thing the filter can do to limit the MSE is to set the result to zero, so we obtain $w_{\text{opt}} = 0$.

(c) Let us define

$$\begin{aligned} p(n) &= \left[\sum_{i=0}^{n} x^2(i)\right]^{-1}, \\ \theta(n) &= \sum_{i=0}^{n} x(i)d(i). \end{aligned}$$

The RLS filter can then be described by the following update equations

$$\begin{aligned} p(n+1) &= p(n) - \frac{p^2(n)x^2(n+1)}{1+p(n)x^2(n+1)} = \frac{p(n)}{1+p(n)x^2(n+1)}, \\ \theta(n+1) &= \theta(n) + x(n+1)d(n+1). \end{aligned}$$

As a result, we obtain

$$\hat{w}(n+1) = p(n+1)\theta(n+1) = \frac{\hat{w}(n) + p(n)x(n+1)d(n+1)}{1+p(n)x^2(n+1)}.$$

(d) If we define

$$p(n) = E\{[\hat{w}(n) - w(n)]^2\}$$

the Kalman gain is given by

$$k(n+1) = p(n)x(n+1)[x^2(n+1)p(n) + 1]^{-1}.$$

As a result, we obtain the following update equation for the Kalman filter

$$\begin{aligned}\hat{w}(n+1) &= \hat{w}(n) + \frac{p(n)x(n+1)}{1 + x^2(n+1)p(n)}[d(n+1) - x(n+1)\hat{w}(n)] \\ &= \frac{\hat{w}(n) + p(n)x(n+1)d(n+1)}{1 + x^2(n+1)p(n)}.\end{aligned}$$

Clearly, if we set the $p(n)$ of the RLS filter equal to the $p(n)$ of the Kalman filter (note that they have different meanings though), we observe that both filters are equivalent. Hence, RLS can be interpreted as a special case of the Kalman filter.

(e) In that case, we have to adapt the state equation as

$$w(n+1) = \frac{1}{\lambda}w(n).$$

This is easy to check in a similar way as above.

Question 4 (12 points)

Consider a signal $z(n)$ consisting of two complex sinusoids in noise:

$$z(n) = A_1 e^{j(\omega_1 n + \phi_1)} + A_2 e^{j(\omega_2 n + \phi_2)} + w(n)$$

where ω_1, ω_2 are the frequencies of the sinusoids, A_1, A_2 are their amplitudes, ϕ_1, ϕ_2 are random phases, considered uniformly distributed in $(0, 2\pi)$, and $w(n)$ is zero mean white noise. We have N samples $n = 0, \cdots, N-1$ and create a periodogram.

(a) What is the power spectrum $P_z(e^{j\omega})$ and the expected value of the periodogram $\hat{P}_z(e^{j\omega})$?

(b) Make a sketch of the periodogram.

(c) How can we estimate the frequencies ω_1, ω_2? How does the finite size of N limit the resolution? What is the effect of the noise? What happens if N becomes very large?

Suppose now that a digital communications signal $x(n)$ is transmitted to a receiver over a channel with two reflection paths. The received sampled signal $y(n)$ consists of two attenuated and delayed copies of $x(n)$, and additive white noise:

$$y(n) = \alpha_1 x(n - n_1) + \alpha_2 x(n - n_2) + w(n),$$

where $\alpha_1 = A_1 e^{j\phi_1}$ and $\alpha_2 = A_2 e^{j\phi_2}$. Here, A_1, A_2 are attenuations, ϕ_1, ϕ_2 are random phase offsets, n_1 and n_2 are the delays of the reflection paths, and $w(n)$ is zero mean complex-valued white noise. The input signal $x(n)$ is a random BPSK signal (i.e., $x(n) \in \{-1, +1\}$).

We would like to estimate the delay of the propagation paths from knowledge of $x(n)$ and $y(n)$.

(d) Let $X(e^{j\omega})$ be the DTFT of $x(n)$, and likewise for $Y(e^{j\omega})$ and $W(e^{j\omega})$.

From the data, we compute

$$Z(e^{j\omega}) := \frac{Y(e^{j\omega})}{X(e^{j\omega})}$$

What is a model for $Z(e^{j\omega})$ in terms of $X(e^{j\omega})$ and $W(e^{j\omega})$?

Hint: Hayes p. 14 has a table of DTFT properties.

(e) If we only have available N samples, we replace the DTFT by the DFT and obtain the sequence $X[k] = X(e^{j\frac{2\pi}{N}k})$ for $k = 0, \cdots, N-1$, and likewise for $Y[k]$ and $W[k]$. From the data, we compute

$$Z[k] := \frac{Y[k]}{X[k]}$$

- Show that an (approximate) model for $Z[k]$ is

$$Z[k] = \alpha_1 e^{-jk\omega_1} + \alpha_2 e^{-jk\omega_2} + E[k]$$

  with $\omega_1 = 2\pi n_1/N$ and $\omega_2 = 2\pi n_2/N$, and $E[k] = \frac{W[k]}{X[k]}$.
- What limits the accuracy of this model?

(f) Is it reasonable to model $W[k]$ as zero mean white noise? Is it reasonable to model $E[k]$ as zero mean white noise? (Motivate your answers.)

(g) Using periodograms, how can we estimate the delays n_1 and n_2 from $Z[k]$, $k = 0, \cdots, N-1$? *Hint: compare to your answers in part (a)-(c).*

How does the finite size of N limit the resolution? What is the effect of the noise? What happens if N becomes very large?

Solution

(a) See Hayes p.401:

$$P_z(e^{j\omega}) = \pi \left[A_1^2 u_0(\omega - \omega_1) + A_2^2 u_0(\omega - \omega_2)\right] + \sigma_n^2$$

$$E\{\hat{P}_z(e^{j\omega})\} = \frac{1}{2}\left[A_1^2 W_B(\omega - \omega_1) + A_2^2 W_B(\omega - \omega_2)\right] + \sigma_n^2$$

where $W_B(\omega) = \frac{1}{N}\left[\frac{\sin(N\omega/2)}{\sin(\omega/2)}\right]^2$.

(b) See Hayes fig 8.7(b). (In the present case, since the signals are complex, the spectrum is not symmetric and runs till $\omega = 2\pi$.)

(c) The peaks in the periodogram correspond to the frequencies. For smaller N, the window function is less sharp (mainlobe width $2\pi/N$).

The noise power raises the spectrum. Also, noise makes the periodogram more noisy (adds variance to it) making it harder to distinguish peaks.

For $N \to \infty$, the peaks become more narrow but the variance remains finite, hence it does not converge to the power spectrum.

(d)

$$Y(e^{j\omega}) = \alpha_1 e^{-j\omega n_1} X(e^{j\omega}) + \alpha_2 e^{-j\omega n_2} X(e^{j\omega}) + W(e^{j\omega})$$

$$Z(e^{j\omega}) = \alpha_1 e^{-j\omega n_1} + \alpha_2 e^{-j\omega n_2} + E(e^{j\omega})$$

with $E(e^{j\omega}) = W(e^{j\omega})/X(e^{j\omega})$.

(e)

$$Z[k] = \alpha_1 e^{-jk\omega_1} + \alpha_2 e^{-jk\omega_2} + E[k]$$

with $\omega_1 = 2\pi n_1/N$ and $\omega_2 = 2\pi n_2/N$.

The relation $X[k] = X(e^{j\frac{2\pi}{N}k})$ is actually an approximation because the DTFT $X(\cdot)$ is based on infinite data and the DFT only on finite data. It is valid if the sequence is zero outside the interval $[0, N-1]$.

Further, the DFT relates to circular convolution, whereas the delay property is valid for linear convolution. Linear convolution is the same as circular convolution if the delayed sequence $y[n]$ is also zero outside the interval $[0, N-1]$. This condition makes sense because we also want $Y[k]$ to relate to $Y(\cdot)$.

Thus, both approximations are edge effects that go away for large N.

(f) Yes: $W[k]$ is related to $w[n]$ via a unitary $N \times N$ matrix $\mathbf{F}$ (the DFT matrix). This transforms a white noise sequence into a white noise sequence. Nearly the same holds for $X[k]$ (although the fact that it is real gives rise to a symmetry.)

Yes, $E[k]$ is a white sequence because it comes from a white sequence where every entry is divided by some number: this does not introduce correlations within the sequence. Also all entries are expected to have the same power.

(g) Take the sequence $Z[k]$ and compute its periodogram. Look for peaks in this plot: they correspond to ω_1 and ω_2.

For finite N, we have 2 problems: the approximation of item (e) doesn't hold, and the finite resolution of the periodogram (as in item (c)). Noise will raise the noise floor in the periodogram and make the estimates noisy (adds a variance). For large N, the resolution becomes better but this variance doesn't go away.