EECS 225A Spring 2005

Selected book problem solutions

Chapters 7, 8, and 9

The following problem solutions should assist you in studying for the second midterm. You should learn more if you give each problem a go before looking at the solution.

7.6 Suppose that a process $x(n)$ has been recorded but there is a missing gap over the interval $[N_1, N_2]$, i.e., $x(n)$ is unknown over this interval.

(a) Derive the optimum estimate of $x(N_1)$ using the data in the semi-infinite interval $(-\infty, N_1 - 1]$.

(b) Derive the optimum estimate of $x(N_1)$ using the data in the semi-infinite interval $[N_2 + 1, \infty)$.

(c) Derive the optimum estimate of $x(N_1)$ that is formed by combining together the two estimates found in parts (a) and (b).

(d) Generalize your result in part (c) to find the optimum estimate of $x(n)$ at an arbitrary point n in the interval $[N_1, N_2]$.

Solution

(a) This is the one-step causal linear prediction problem. The optimum estimate of $x(N_1)$ is

$$\widehat{x}_1(N_1) = \sum_{k=0}^{\infty} h_1(k)x(N_1 - 1 - k)$$

where

$$H_1(z) = \frac{1}{Q(z)}\Big[zQ(z)\Big]_+$$

and $Q(z)$ is the minimum phase factor of $P_x(z)$,

$$P_x(z) = \sigma_0^2 Q(z)Q(z^{-1})$$

(b) Here we want to design the optimum filter for predicting *backwards* $N_2 - N_1 + 1$ samples. The system function for the optimum causal $(N_2 - N_1 + 1)$-step *forward* predictor is

$$H_2(z) = \frac{1}{Q(z)}\Big[z^{N_2 - N_1 + 1}Q(z)\Big]_+$$

Therefore, the optimum estimate of $x(N_1)$ is

$$\widehat{x}_2(N_1) = \sum_{k=0}^{\infty} h_2(k)x(N_2 + 1 + k)$$

(c) We are given two estimates of $x(N_1)$, which we have called $\widehat{x}_1(N_1)$ and $\widehat{x}_2(N_1)$. Let the variance of these estimates be denoted by σ_1^2 and σ_2^2, respectively. We would now like to find the optimum estimate of $x(N_1)$ using an estimate of the form

$$\widehat{x}(N_1) = K\widehat{x}_1(N_1) + (1 - K)\widehat{x}_2(N_1)$$

where K is a constant that is to be determined. Note that this form for the estimate guarantees that $\widehat{x}(N_1)$ will be unbiased if $\widehat{x}_1(N_1)$ and $\widehat{x}_2(N_1)$ are unbiased. To find the value of K that minimizes the mean-square error,

$$\xi = E\Big\{\big[x(N_1) - \widehat{x}(N_1)\big]^2\Big\}$$

we set the derivative of ξ with respect to K equal to zero and solve. With

$$e(N_1) = x(N_1) - \widehat{x}(N_1) = K\big[x(N_1) - \widehat{x}_1(N_1)\big] + (1 - K)\big[x(N_1) - \widehat{x}_2(N_1)\big]$$

we have

$$\frac{\partial \xi}{\partial K} = 2E\Big\{e(N_1)\Big(\big[x(N_1) - \widehat{x}_1(N_1)\big] - \big[x(N_1) - \widehat{x}_2(N_1)\big]\Big)\Big\} = 0$$

Substituting for $e(N_1)$ and taking the expected value we have, assuming that the estimation errors $e_1(N_1) = x(N_1) - \widehat{x}_1(N_1)$ and $e_2(N_1) = x(N_1) - \widehat{x}_2(N_1)$ are uncorrelated,

$$K\sigma_1^2 - (1-K)\sigma_2^2 = 0$$

Solving for K we have

$$K = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

which leads to the following estimate for $x(N_1)$,

$$\widehat{x}(N_1) = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\widehat{x}_1(N_1) + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\widehat{x}_2(N_1)$$

(d) For an arbitrary point N_0 in the interval $[N_1, N_2]$, the result derived in part (c) produces the optimum estimate of $x(N_0)$ provided $H_1(z)$ is replaced with

$$H_1(z) = \frac{1}{Q(z)}\left[z^{N_0 - N_1 + 1}Q(z)\right]_+$$

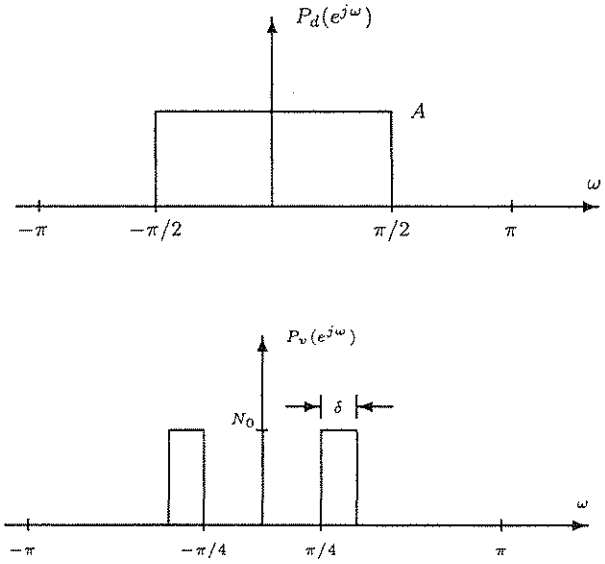
and $H_2(z)$ is replaced with

$$H_2(z) = \frac{1}{Q(z)}\left[z^{N_2 - N_0 + 1}Q(z)\right]_+$$

7.8 Suppose that we would like to estimate a signal $d(n)$ from the noisy observations

$$x(n) = d(n) + v(n)$$

where the noise, $v(n)$, is uncorrelated with $d(n)$. The power spectral densities of $d(n)$ and $v(n)$ are shown in the following figure.



(a) Design a *noncausal* Wiener smoothing filter for estimating $d(n)$ from $x(n)$,

$$\widehat{d}(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k)$$

(b) Compute the mean-square error $E\{|d(n) - \widehat{d}(n)|^2\}$ and compare it to the mean-square error that results when $h(n) = \delta(n)$, i.e., with no filtering of $x(n)$.

Solution

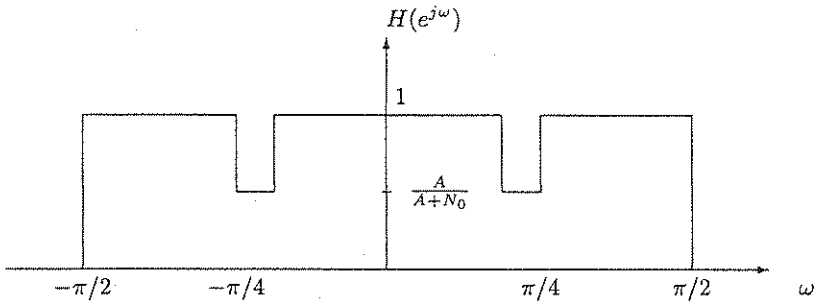
(a) The frequency response of the optimum noncausal Wiener filter is

$$H(e^{j\omega}) = \frac{P_{dx}(e^{j\omega})}{P_x(e^{j\omega})} = \frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_w(e^{j\omega})}$$

Using the given power spectral densities for $d(n)$ and $w(n)$, it follows that the frequency response of the Wiener filter is as shown in the figure below.

(b) The minimum mean-square error is

$$\begin{aligned} \xi_{min} &= r_d(0) - \sum_{l=-\infty}^{\infty} h(l) r_{dx}(l) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[P_d(e^{j\omega}) - H(e^{j\omega}) P_d(e^{-j\omega}) \right] d\omega \end{aligned}$$



Evaluating the integral we find

$$\xi_{min} = \frac{\delta A}{\pi} \frac{N_0}{N_0 + A}$$

The mean-square error using $h(n) = \delta(n)$ is

$$\xi = E\left[v^2(n)\right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_v(e^{j\omega}) d\omega = \frac{\delta N_0}{\pi}$$

7.11 In this problem we consider the design of an optimum smoothing filter for estimating a process $d(n)$ from the measurements

$$x(n) = d(n) + v(n)$$

Our goal is to use a noncausal FIR filter that has a system function of the form:

$$W(z) = \sum_{k=-p}^{p} w(k) z^{-k}$$

In other words, we want to produce an estimate of $d(n)$ as follows

$$\widehat{d}(n) = \sum_{k=-p}^{p} w(k) x(n-k)$$

(a) Derive the Wiener-Hopf equations that define the set of coefficients that minimize the mean-square error

$$\xi = E\{ |d(n) - \widehat{d}(n)|^2 \}$$

(b) How would the Wiener-Hopf equations derived in part (a) change if we used a causal filter with the same number of coefficients? In other words, if the system function was of the form

$$W(z) = \sum_{k=0}^{2p} w(k) z^{-k}$$

how would you modify your equations in (a)?

(c) State qualitatively when you might prefer the noncausal filter over the causal filter and vice versa. For example, for what types of signals and for what types of noise would you expect a causal filter to be superior to the noncausal filter?

(d) FIR digital filters with linear phase (or zero phase) are important in many signal processing applications where frequency dispersion due to nonlinear phase is harmful. An FIR filter with zero phase is characterized by the property that

$$w(n) = w(-n)$$

Thus, the system function may be written as

$$W(z) = w(0) + \sum_{k=1}^{p} w(k)[z^{-k} + z^{k}]$$

Derive the Wiener-Hopf equations that define the optimum zero phase smoothing filter.

(e) With $r_d(k) = 4(0.5)^{|k|}$ and $r_v(k) = \delta(k)$, find the optimum values for the filter coefficients $w(0)$ and $w(1)$ in the zero phase filter

$$W(z) = w(0) + w(1)\left[z + z^{-1}\right]$$

Solution

(a) Using a noncausal filter of the form

$$W(z) = \sum_{k=-p}^{p} w(k) z^{-k}$$

the estimate of $d(n)$ is given by

$$\widehat{d}(n) = \sum_{k=-p}^{p} w(k) x(n-k)$$

With

$$\xi = E\left\{|e(n)|^2\right\} = E\left\{|d(n) - \widehat{d}(n)|^2\right\}$$

the coefficients that minimize the mean-square error are solutions to the following equations

$$\frac{\partial \xi}{\partial w^*(k)} = E\left\{e(n)\frac{\partial e^*(n)}{\partial a^*(k)}\right\} = 0 \quad ; \quad k = -p, \ldots, p$$

Since

$$e(n) = d(n) - \sum_{l=-p}^{p} w(l) x(n-l)$$

then

$$\frac{\partial e^*(n)}{\partial w^*(k)} = -x^*(n-k)$$

and

$$\frac{\partial \xi}{\partial w^*(k)} = -E\{e(n)x^*(n-k)\} = 0 \quad ; \quad k = -p, \ldots, p$$

Using the definition of $e(n)$ this becomes

$$E\{d(n)x^*(n-k)\} - \sum_{l=-p}^{p} w(l) E\{x(n-l)x^*(n-k)\} = 0$$

and, with $r_x(k) = E\{x(n)x^*(n-k)\}$ and $r_{dx}(k) = E\{d(n)x^*(n-k)\}$ we have

$$\sum_{l=-p}^{p} w(l) r_x(k-l) = r_{dx}(k) \quad ; \quad k = -p, \ldots, p$$

which is a set of $(2p+1)$ linear equations in the $(2p+1)$ unknowns $w(k)$, $k = -p, \ldots, p$. These equations may be written in matrix form as follows

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

We may now evaluate the mean-square error,

$$\begin{aligned} \xi &= E\{|e(n)|^2\} = E\left\{ e(n)\Big[d(n) - \sum_{l=-p}^{p} w(l)x(n-l)\Big]^* \right\} \\ &= E\{e(n)d^*(n)\} - \sum_{l=-p}^{p} w(l) E\{e(n)x^*(n-l)\} \end{aligned}$$

However, from the orthogonality condition, $E\{e(n)x^*(n-k)\} = 0$, it follows that the second term is zero, and we have

$$\xi_{\min} = E\{e(n)d^*(n)\} = E\left\{ \Big[d(n) - \sum_{l=-p}^{p} w(l)x(n-l)\Big] d^*(n) \right\}$$

Finally, taking expected values we find

$$\xi_{\min} = r_d(0) - \sum_{l=-p}^{p} w(l) r_{dx}^*(l)$$

(b) If we change the filter in (a) to a causal smoothing filter of the form

$$W(z) = \sum_{k=0}^{2p} w(k) z^{-k}$$

then the only thing that changes in the derivations above is the limits on the summations. In particular, the Wiener-Hopf equations would take the form

$$\sum_{l=0}^{2p} w(l) r_x(k-l) = r_{dx}^*(k) \quad ; \quad k = 0, \ldots, 2p$$

and the minimum error would be

$$\xi_{\min} = r_d(0) - \sum_{l=0}^{2p} w(l) r_{dx}^*(l)$$

(c) If the autocorrelation sequence for $d(n)$ is generally larger over the interval $[-p,\ldots,p]$ than it is over the interval $[0,\ldots,2p]$, then the zero phase smoothing filter would be better to use. Otherwise, the causal filter would be better. For example, with a signal that has an autocorrelation sequence $\mathbf{r}_d = [1,0,0,0,.9,\ldots]$, you may wish to compare the performance of the optimum causal and noncausal filters when $p = 2$.

(d) Using a zero phase smoothing filter

$$W(z) = w(0) + \sum_{k=1}^{p} w(k)\left[z^{-k} + z^{k}\right]$$

the estimate of $d(n)$ is now given by

$$\widehat{d}(n) = w(0)x(n) + \sum_{l=1}^{p} w(l)\big[x(n-l) + x(n+l)\big]$$

Again, with

$$\xi = E\big\{|e(n)|^2\big\} = E\big\{|d(n) - \widehat{d}(n)|^2\big\}$$

we differentiate ξ with respect to $w^*(k)$ as follows

$$\frac{\partial \xi}{\partial w^*(k)} = E\left\{e(n)\frac{\partial e^*(n)}{\partial w^*(k)}\right\} = 0 \quad ; \quad k = 0,\ldots,p$$

However, since

$$\frac{\partial e^*(n)}{\partial w^*(0)} = -x^*(n)$$

and

$$\frac{\partial e^*(n)}{\partial w^*(k)} = -x^*(n-k) - x^*(n+k) \quad ; \quad k = 1,\ldots,p$$

then we have

$$E\big\{e(n)x^*(n)\big\} = 0$$

and

$$E\big\{e(n)[x(n-k) + x(n+k)]^*\big\} = 0 \quad ; \quad k = 1,\ldots,p$$

If we now substitute for $e(n)$ we find

$$r_{dx}(0) - \left\{w(0)r_x(0) + \sum_{l=1}^{p} w(l)\big[r_x(l) + r_x(-l)\big]\right\} = 0$$

and

$$2r_{dx}(k) - \left\{2w(0)r_x(k) + \sum_{l=1}^{p} 2w(l)\big[r_x(k-l) + r_x(k+l)\big]\right\} = 0 \quad ; \quad k = 1,\ldots,p$$

Thus,

$$w(0)r_x(0) + \sum_{k=1}^{p} 2w(k)r_x(k) = r_{dx}(0)$$

and

$$w(0)r_x(k) + \sum_{l=1}^{p} w(l)\big[r_x(k-l) + r_x(k+l)\big] = r_{dx}(k) \quad ; \quad k = 1,\ldots,p$$

(e) The coefficients for the third-order zero phase filter are the solution to the equations,

$$\begin{bmatrix} r_x(0) & 2r_x(1) \\ r_x(1) & r_x(0) + r_x(2) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} r_{dx}(0) \\ r_{dx}(1) \end{bmatrix}$$

Incorporating the given values for the autocorrelations, we have

$$\begin{bmatrix} 5 & 4 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

and the solution is

$$W(z) = \frac{8}{11} + \frac{1}{11}\left[z^{-1} + z\right]$$

8.2 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 available data in a radix-2 FFT algorithm, and it is required that the estimate have a resolution of at least 10 Hz. Suppose that we use Bartlett's method of periodogram averaging.

(a) If the data is sampled at the Nyquist rate, what is the minimum section length that you may use to get the desired resolution?

(b) Using the minimum section length determined in part (a), with 10 seconds of data, how many sections are available for averaging?

(c) How does your choice of the sampling rate affect the resolution and variance of your estimate? Are there any benefits to sampling above the Nyquist rate?

Solution

(a) If we sample at the Nyquist rate, $f_s = 10$kHz, then 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

$$\text{Res}[\widehat{P}_{PER}(\omega)] = \Delta\omega = 0.89\frac{2\pi}{M}$$

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

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

(b) Sampling at 10 kHz, 10 seconds of data corresponds to $N = (10)(10 \times 10^3) = \times 10^5$ samples. Therefore, with a 1024-point DFT the number of sections we may have in Bartlett's method is

$$K = [N/1024] = 98$$

(c) 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 within a T second interval. Therefore, since the variance (normalized) is

$$V = M/N$$

increasing the sampling rate has no effect. Thus, provided that the sampling rate is not less than the Nyquist frequency, the resolution does not depend on the sampling rate.

8.5 Many commercial *Fourier analyzers* continuously update the estimate of the power spectrum of a process $x(n)$ by exponential averaging periodograms as follows,

$$\widehat{P}_i(e^{j\omega}) = \alpha \widehat{P}_{i-1}(e^{j\omega}) + \frac{1-\alpha}{N}\left|\sum_{n=0}^{N-1} x_i(n)e^{-jn\omega}\right|^2$$

where $x_i(n) = x(n+Ni)$ is the ith sequence of N data values. This update equation is initialized with $\widehat{P}_{-1}(e^{j\omega}) = 0$.

(a) Qualitatively describe the philosophy behind this method, and discuss how the value for the weighting factor α should be selected.

(b) Assuming that successive periodograms are uncorrelated and that $0 < \alpha < 1$, find the mean and variance of $\widehat{P}_i(e^{j\omega})$ for a Gaussian random process.

(c) Repeat the analysis in part (b) if the periodograms are replaced with modified periodograms.

Solution

(a) As data is being read by a spectrum analyzer, the goal is to continuously update the estimate. As each data record of length N is collected, the periodogram is computed, and *averaged* with the previous spectrum estimate. Althogh a running average could be formed, this would assume that the process is stationary. Selecting a value of $0 < \alpha < 1$ allows the estimate to *forget* $\widehat{P}_i(e^{j\omega})$ as more data is collected. In the extreme case in which $\alpha = 0$, $\widehat{P}_i(e^{j\omega})$ is the periodogram of the most recent N data values. As we will see in part (b), $\widehat{P}_i(e^{j\omega})$ is an exponentially weighted average of the previous periodograms.

(b) If we define

$$Q_i(e^{j\omega}) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x_i(n)e^{-jn\omega}\right|^2$$

then the expression for the ith spectrum estimate, $P_i(e^{j\omega})$ is

$$\widehat{P}_i(e^{j\omega}) = \alpha\widehat{P}_{i-1}(e^{j\omega}) + \frac{1-\alpha}{N}Q_i(e^{j\omega})$$

which is a difference equation for $\widehat{P}_i(e^{j\omega})$. Since the initial conditions are zero, $\widehat{P}_{-1}(e^{j\omega}) = 0$, then the solution for $\widehat{P}_i(e^{j\omega})$ is

$$\widehat{P}_i(e^{j\omega}) = \sum_{k=0}^{i}(1-\alpha)\alpha^k Q_k(e^{j\omega})$$

Taking the expected value we have

$$E\{\widehat{P}_i(e^{j\omega})\} = \sum_{k=0}^{i}(1-\alpha)\alpha^k E\{Q_k(e^{j\omega})\}$$

Since $Q_k(e^{j\omega})$ is the periodogram of $x_k(n)$, then

$$E\{Q_k(e^{j\omega})\} = \frac{1}{2\pi}P_x(e^{j\omega}) * W_B(e^{j\omega})$$

and

$$\begin{aligned}
E\{\widehat{P}_i(e^{j\omega})\} &= \frac{1}{2\pi}\sum_{k=0}^{i}(1-\alpha)\alpha^k\left[P_x(e^{j\omega}) * W_B(e^{j\omega})\right] \\
&= \frac{1}{2\pi}\left[P_x(e^{j\omega}) * W_B(e^{j\omega})\right](1-\alpha)\sum_{k=0}^{p}\alpha^k \\
&= \frac{1}{2\pi}\left[P_x(e^{j\omega}) * W_B(e^{j\omega})\right](1-\alpha)\frac{1-\alpha^{i+1}}{1-\alpha} \\
&= (1-\alpha^{i+1})\frac{1}{2\pi}\left[P_x(e^{j\omega}) * W_B(e^{j\omega})\right]
\end{aligned}$$

For the variance, we proceed in the same way, using the fact that the variance of the periodogram is

$$\text{var}\{\widehat{P}_{\text{per}}(e^{j\omega})\} \approx P_x^2(e^{j\omega})$$

The result is

$$\text{var}\{\widehat{P}_i(e^{j\omega})\} \approx (1-\alpha^{i+1})\frac{1}{2\pi}P_x^2(e^{j\omega})$$

(c) For modified periodograms, the only change that is necessary is to use

$$E\{Q_i(e^{j\omega})\} = \frac{1}{NU}\left|P_x(e^{j\omega}) * W_B(e^{j\omega})\right|^2$$

where

$$U = \frac{1}{N}\sum_{n=0}^{N-1}|w(n)|^2$$

Substituting this into the expression in part (b) for

$$E\{\widehat{P}_i(e^{j\omega})\} = (1-\alpha^{i+1})\frac{1}{2\pi NU}\left|P_x(e^{j\omega}) * W_B(e^{j\omega})\right|^2$$

The variance is the same as for the periodogram.

8.26 A random process may be classified in terms of the properties of the prediction error sequence ϵ_k that is produced when fitting an all-pole model to the process. Listed below are five different classifications for the error sequence:

1. $\epsilon_k = c > 0$ for all $k \geq 0$.
2. $\epsilon_k = c > 0$ for all $k \geq k_0$ for some $k_0 > 0$.
3. $\epsilon_k \rightarrow c$ as $k \rightarrow \infty$ where $c > 0$.
4. $\epsilon_k \rightarrow 0$ as $k \rightarrow \infty$.
5. $\epsilon_k = 0$ for all $k \geq k_0$ for some $k_0 > 0$.

For each of these classifications, describe as completely as possible the characteristics that may be attributed to the process and its power spectrum.

Solution

1. If $\epsilon_k = c > 0$ for all $k \geq 0$, then the reflection coefficients are zero, $\Gamma_k = 0$ for all k. Therefore, the process is not predictable, which is consistent with a white noise process with a power spectrum that is constant, $P_x(e^{j\omega}) = c$.
2. If $\epsilon_k = c > 0$ for all $k \geq k_0$ for some $k_0 > 0$, then the reflection coefficients Γ_k are equal to zero for all $k \geq k_0$. This is consistent with a finite-order autoregressive process of order k_0. Only a finite past history is used to model the process.
3. If $\epsilon_k \rightarrow c$ as $k \rightarrow \infty$ where $c > 0$, then the entire past history of the process is useful in predicting (modeling) the process. Since ϵ_k approaches a limit that is not zero, this behavior is consistent with a moving average or autoregressive moving average process.

4. If $\epsilon_k \to 0$ as $k \to \infty$, the process is perfectly predictable given the infinite past history of the process. Recall that the prediction error, expressed in terms of the power spectrum, is (see Eq. (7.72), the Kolmogorov-Szegö formula)

$$\lim_{k\to\infty} \epsilon_k = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln P_x(e^{j\omega})d\omega\right\}$$

This will be zero only if $P_x(e^{j\omega}) = 0$ over some finite interval. This behavior is consistent with a lowpass, bandpass, or highpass process in which $P_x(e^{j\omega}) = 0$ for $\omega \in [\omega_1, \omega_2]$.

5. If $\epsilon_k = 0$ for all $k \geq k_0$ for some $k_0 > 0$, then the process is perfectly predictable from the past history of $x(n)$ that is finite in length. This corresponds to the case in which $\Gamma_{k_0} = \pm 1$ and, therefore, represents a harmonic process (line spectrum).

9.3 Newton's method is an iterative algorithm that may be used to find the minimum of a nonlinear function. Applied to the minimization of the mean-square error

$$\xi(n) = E\{e^2(n)\}$$

where $e(n) = d(n) - \mathbf{w}^T\mathbf{x}(n)$, Newton's method is

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \tfrac{1}{2}\mathbf{R}_x^{-1}\nabla\xi(n)$$

where $\mathbf{R}_x$ is the autocorrelation matrix of $x(n)$. Introducing a step-size parameter μ, Newton's method becomes

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \tfrac{1}{2}\mu\mathbf{R}_x^{-1}\nabla\xi(n)$$

Comparing this to the steepest descent algorithm, we see that the step size μ is replaced with a matrix, $\mu\mathbf{R}_x^{-1}$, which alters the descent direction.

(a) For what values of μ is Newton's method stable, i.e., for what values of μ will $\mathbf{w}_n$ converge?

(b) What is the optimum value of μ, i.e., for what value of μ is the convergence the fastest?

(c) Suppose that we form an LMS version of Newton's method by replacing the gradient with a gradient estimate

$$\widehat{\nabla}\xi(n) = \nabla e^2(n)$$

Derive the coefficient update equation that results from using this gradient estimate and describe how it differs from the LMS algorithm.

(d) Derive an expression that describes the time evolution of $E\{\mathbf{w}_n\}$ using the LMS Newton algorithm derived in part (c).

Solution

(a) Evaluating the gradient vector we have

$$\nabla\xi(n) = 2E\{e(n)\nabla e(n)\} = -2E\{e(n)\mathbf{x}(n)\} = -2\mathbf{r}_{dx} + 2\mathbf{R}_x\mathbf{w}_n$$

Thus,

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \tfrac{1}{2}\mu\mathbf{R}_x^{-1}\left[2\mathbf{R}_x\mathbf{w}_n - 2\mathbf{r}_{dx}\right]$$

and we have

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu\mathbf{w}_n + \mu\mathbf{w} = (1-\mu)\mathbf{w}_n + \mu\mathbf{w}$$

Thus, the Newton algorithm is stable for $0 < \mu < 2$.

(b) The convergence is the fastest when $\mu = 1$. Note, in fact, that when $\mu = 1$, the Newton iteration converges in one step to $\mathbf{w}$.

(c) The gradient approximation is

$$\nabla e^2(n) = -2e(n)\mathbf{x}(n)$$

Therefore the LMS-type algorithm is

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu e(n)\mathbf{R}_x^{-1}\mathbf{x}(n)$$

Comparing this to the LMS algorithm we see that the step direction is changed from $\mathbf{x}(n)$ to $\mathbf{R}^{-1}\mathbf{x}(n)$.

(d) From (c) we see that

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu\mathbf{R}_x^{-1}\mathbf{x}(n)d(n) - \mu\mathbf{R}_x^{-1}\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{w}_n$$

Assuming that $\mathbf{x}(n)$ is uncorrelated with the filter tap weight vector, $\mathbf{w}_n$, then

$$E\{\mathbf{w}_{n+1}\} = E\{\mathbf{w}_n\} + \mu\mathbf{R}_x^{-1}\mathbf{r}_{dx} - \mu\mathbf{R}_x^{-1}\mathbf{R}_x E\{\mathbf{w}_n\}$$

which becomes

$$E\{\mathbf{w}_{n+1}\} = (1-\mu)E\{\mathbf{w}_n\} + \mu\mathbf{w}$$

9.4 One way to derive the steepest descent algorithm for solving the normal equations $\mathbf{R}_x\mathbf{w} = \mathbf{r}_{dx}$ is to use a power series expansion for the inverse of $\mathbf{R}_x$. This expansion is

$$\mathbf{R}_x^{-1} = \mu \sum_{k=0}^{\infty} (\mathbf{I} - \mu\mathbf{R}_x)^k$$

where $\mathbf{I}$ is the identity matrix and μ is a positive constant. In order for this expansion to converge, $\mathbf{R}_x$ must be positive definite and the constant μ must lie in the range

$$0 < \mu < 2/\lambda_{\max}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}_x$.

(a) Let

$$\mathbf{R}_x^{-1}(n) = \mu \sum_{k=0}^{n} (\mathbf{I} - \mu\mathbf{R}_x)^k$$

be the nth-order approximation to $\mathbf{R}_x^{-1}$, and let

$$\mathbf{w}_n = \mathbf{R}_x^{-1}(n)\mathbf{r}_{dx}$$

be the nth-order approximation to the desired solution $\mathbf{w} = \mathbf{R}_x^{-1}\mathbf{r}_{dx}$. Express $\mathbf{R}_x^{-1}(n+1)$ in terms of $\mathbf{R}_x^{-1}(n)$, and show how this may be used to derive the steepest descent algorithm

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu\big[\mathbf{R}_x\mathbf{w}_n - \mathbf{r}_{dx}\big]$$

(b) If the statistics of $x(n)$ are unknown, then $\mathbf{R}_x$ is unknown and the expansion for $\mathbf{R}_x^{-1}$ in part (a) cannot be evaluated. However, suppose that we approximate $\mathbf{R}_x = E\{\mathbf{x}(n)\mathbf{x}^T(n)\}$ at time n as follows

$$\widehat{\mathbf{R}}_x(n) = \mathbf{x}(n)\mathbf{x}^T(n)$$

and use, as the nth-order approximation to $\mathbf{R}_x^{-1}$,

$$\widehat{\mathbf{R}}_x^{-1}(n) = \mu \sum_{k=0}^{n} \Big[\mathbf{I} - \mu\mathbf{x}(k)\mathbf{x}^T(k)\Big]^k$$

Express $\widehat{\mathbf{R}}_x^{-1}(n+1)$ in terms of $\widehat{\mathbf{R}}_x^{-1}(n)$ and use this expression to derive a recursion for $\mathbf{w}_n$.

(c) Compare your recursion derived in part (b) to the LMS algorithm.

Solution

(a) Using the nth-order approximation to $\mathbf{R}_x^{-1}$,

$$\mathbf{R}_x^{-1}(n) = \mu \sum_{k=0}^{n} \big(\mathbf{I} - \mu\mathbf{R}_x\big)^k$$

we have

$$\mathbf{R}_x^{-1}(n+1) = \mu \sum_{k=0}^{n+1} \big(\mathbf{I} - \mu\mathbf{R}_x\big)^k = \mu\big(\mathbf{I} - \mu\mathbf{R}_x\big) \sum_{k=0}^{n} \big(\mathbf{I} - \mu\mathbf{R}_x\big)^k + \mu\mathbf{I}$$

Therefore,

$$\mathbf{R}_x^{-1}(n+1) = \big(\mathbf{I} - \mu\mathbf{R}_x\big)\mathbf{R}_x^{-1}(n) + \mu\mathbf{I}$$

Multiplying both sides of the equation by $\mathbf{r}_{dx}$ on the right we have

$$\mathbf{w}_{n+1} = \big(\mathbf{I} - \mu\mathbf{R}_x\big)\mathbf{w}_n + \mu\mathbf{r}_{dx}$$

which is the steepest descent algorithm.

(b) Using the approximation $\widehat{\mathbf{R}}_x = \mathbf{x}(n)\mathbf{x}^T(n)$ for $\mathbf{R}_x$ we have

$$\widehat{\mathbf{R}}_x^{-1}(n+1) = \big[\mathbf{I} - \mu\mathbf{x}(n)\mathbf{x}^T(n)\big]\widehat{\mathbf{R}}_x^{-1}(n) + \mu\mathbf{I}$$

Multiplying both sides of the equation by $\mathbf{r}_{dx}$ on the right, we have

$$\mathbf{w}_{n+1} = \big[\mathbf{I} - \mu\mathbf{x}(n)\mathbf{x}^T(n)\big]\mathbf{w}_n + \mu\mathbf{r}_{dx}$$

(c) The recursion in (b) is the same as the p-vector algorithm (see Problem 14). However, if we use the approximation

$$\widehat{\mathbf{r}}_{dx} = d(n)\mathbf{x}(n)$$

then the recursion becomes equivalent to the LMS algorithm.

9.7 Suppose that the input to an FIR LMS adaptive filter is a first-order autoregressive process with an autocorrelation

$$r_x(k) = c\alpha^{|k|}$$

where $c > 0$ and $0 < \alpha < 1$. Suppose that the step size μ is

$$\mu = \frac{1}{5\lambda_{\max}}$$

(a) How does the rate of convergence of the LMS algorithm depend upon the value of α?

(b) What effect does the value of c have on the rate of convergence?

(c) How does the rate of convergence of the LMS algorithm depend upon the desired signal $d(n)$?

Solution

(a) Recall that $\lambda_{\max}$ and $\lambda_{\min}$ are bounded by the power spectrum as follows,

$$\lambda_{\max} \leq \max_{\omega}\left[P_x(e^{j\omega})\right] = c\frac{1+\alpha}{1-\alpha}$$

$$\lambda_{\min} \geq \min_{\omega}\left[P_x(e^{j\omega})\right] = c\frac{1-\alpha}{1+\alpha}$$

Since the time constant for convergence is proportional to the condition number and, for large p,

$$\chi = \frac{\lambda_{\max}}{\lambda_{\min}} \approx \left(\frac{1+\alpha}{1-\alpha}\right)^2$$

then, as α increases, τ increases, and the convergence is slower.

(b) As the constant c changes, the eigenvalues are scaled by c. However, the condition number χ is unaffected. Therefore, c does not affect the time constant for convergence.

(c) The desired signal $d(n)$ has no effect on the rate of convergence.

9.8 The coefficient update equation for the LMS adaptive filter applies a correction to the weight $\mathbf{w}_n$ at time n as follows

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu e(n)\mathbf{x}^*(n)$$

The *Block LMS* algorithm, on the other hand, accumulates these corrections for L samples, beginning at time n, while holding the weight vector $\mathbf{w}_n$ constant. A correction is then applied at the end of the block to form an update at time $n+L$ as follows

$$\mathbf{w}_{n+L} = \mathbf{w}_n + \mu \sum_{l=0}^{L-1} e(n+l)\mathbf{x}^*(n+l)$$

where

$$e(n+l) = d(n+l) - \mathbf{w}_n^T \mathbf{x}(n+l)$$

for $l = 0, 1, \ldots, L-1$.

(a) By evaluating the behavior of $E\{\mathbf{w}_n\}$ as a function of n, determine the conditions on the step size μ that are necessary for the block LMS algorithm to converge in the mean.

(b) Discuss the advantages and/or disadvantages of the block LMS algorithm compared to the standard LMS algorithm.

Solution

(a) The expected value of the weight vector $\mathbf{w}_{n+L}$ is

$$\begin{aligned} E\{\mathbf{w}_{n+L}\} &= E\{\mathbf{w}_n\} + \mu \sum_{l=0}^{L-1} E\{e(n+l)\mathbf{x}^*(n+l)\} \\ &= E\{\mathbf{w}_n\} + \mu \sum_{l=0}^{L-1} E\{[d(n+l) - \mathbf{w}_n^T \mathbf{x}(n+l)]\mathbf{x}^*(n+l)\} \end{aligned}$$

With the independence assumption, this becomes

$$\begin{aligned} E\{\mathbf{w}_{n+L}\} &= \left[\mathbf{I} - \mu \sum_{l=0}^{L-1} E\{\mathbf{x}^T(n+l)\mathbf{x}(n+l)\}\right] E\{\mathbf{w}_n\} + \mu \sum_{l=0}^{L-1} E\{d(n+1)\mathbf{x}(n+l)\} \\ &= (\mathbf{I} - \mu L \mathbf{R}_x) E\{\mathbf{w}_n\} + \mu L \mathbf{r}_{dx} \end{aligned}$$

Therefore, for convergence in the mean it is necessary for the step size μ to satisfy the inequality

$$0 < \mu < \frac{2}{L\lambda_{\max}}$$

(b) The advantage of the block LMS algorithm is that it produces a more accurate gradient estimate (due to the averaging), and a lower misadjustment. The disadvantage, however, is that it is more difficult to track rapidly varying processes.

9.11 The LMS adaptive filter minimizes the instantaneous squared error

$$\xi(n) = |e(n)|^2$$

Consider the modified functional

$$\xi'(n) = |e(n)|^2 + \beta \mathbf{w}_n^H \mathbf{w}_n$$

where $\beta > 0$.

(a) Derive the LMS coefficient update equation for $\mathbf{w}_n$ that minimizes $\xi'(n)$.

(b) Determine the condition on the step size μ that will ensure that $\mathbf{w}_n$ converges in the mean.

(c) If μ is small enough so that $\mathbf{w}_n$ converges in the mean, what does $\mathbf{w}_n$ converge to?

Solution

(a) First we evaluate the gradient of $\xi'(n)$,

$$\nabla \xi'(n) = \nabla |e(n)|^2 + \beta \nabla \left[\mathbf{w}_n^H \mathbf{w}_n \right]$$

The first term is the same as we have for the LMS algorithm,

$$\nabla |e(n)|^2 = -e(n) x^*(n)$$

whereas the second term is equal to

$$\nabla \left[\mathbf{w}_n^H \mathbf{w}_n \right] = \mathbf{w}_n$$

Therefore, the LMS update equation is

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu e(n) \mathbf{x}^*(n) - \mu \beta \mathbf{w}_n = (1 - \mu\beta)\mathbf{w}_n + \mu e(n) \mathbf{x}^*(n)$$

(b) Repeating the analysis that was done for the LMS adaptive filter we have

$$E\{\mathbf{w}_{n+1}\} = \left[(1 - \mu\beta)\mathbf{I} - \mu \mathbf{R}_x \right] E\{\mathbf{w}_n\} + \mu \mathbf{r}_{dx}$$

Therefore, for stability (convergence in the mean) we require

$$\left| (1 - \beta\mu) - \mu\lambda_k \right| < 1 \quad ; \quad k = 0, 1, 2, \ldots, p$$

which implies that

$$0 < \mu < \frac{2}{\beta + \lambda_{\max}}$$

(c) If $\mathbf{w}_n$ converges in the mean to, say $\mathbf{w}_\infty$, then

$$\mathbf{w}_\infty = (1 - \beta\mu)\mathbf{w}_\infty - \mu \mathbf{R}_x \mathbf{w}_\infty + \mu \mathbf{r}_{dx}$$

This implies that

$$\beta\mu \mathbf{w}_\infty + \mu \mathbf{R}_x \mathbf{w}_\infty = \mu \mathbf{r}_{dx}$$

or

$$\left(\mathbf{R}_x + \beta \mathbf{I} \right) \mathbf{w}_\infty = \mathbf{r}_{dx}$$

Therefore,

$$\mathbf{w}_\infty = \left(\mathbf{R}_x + \beta \mathbf{I} \right)^{-1} \mathbf{r}_{dx}$$

9.20 In some applications, it may be necessary to delay the update of the filter coefficients for a short period of time. For example, in decision-directed feedback equalization, if a sophisticated algorithm such a Viterbi decoder is used to improve the decisions, then the desired signal and thus the error is not available until a number of samples later. Therefore, assume that $x(n)$ is real-valued, and consider the delayed LMS algorithm that has a filter coefficient update equation given by

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu e(n - n_0)\mathbf{x}(n - n_0)$$

where

$$e(n - n_0) = d(n - n_0) - y(n - n_0)$$

Note that if the delay, n_0, is equal to zero then we have the conventional LMS algorithm.

(a) For $n_0 = 1$, determine the values of μ for which the delayed LMS algorithm converges in the mean.

(b) If $\lambda_k = 1$, for $k = 1, \ldots, p$ and if the step size $\mu = 0.1$, find the time constant, τ_0 for the LMS adaptive filter ($n_0 = 0$) and the time constant τ_1 for the delayed LMS adaptive filter with $n_0 = 1$.

Solution

(a) With $n_0 = 1$ the delayed LMS adaptive filter update equation is

$$\begin{aligned} \mathbf{w}_{n+1} &= \mathbf{w}_n + \mu e(n-1)\mathbf{x}(n-1) \\ &= \mathbf{w}_n + \mu \left[d(n-1) - \mathbf{w}_{n-1}^T \mathbf{x}(n-1)\right] \mathbf{x}(n-1) \end{aligned}$$

Taking the expected value, assuming that the weight vector $\mathbf{w}_n$ is uncorrelated with the data vector $\mathbf{x}(n)$, we have

$$E\{\mathbf{w}_{n+1}\} = E\{\mathbf{w}_n\} - \mu \mathbf{R}_x E\{\mathbf{w}_{n-1}\} + \mu \mathbf{r}_{dx}$$

Thus, the expected value of the weight vector satisfies a second-order difference equation. Diagonalizing the autocorrelation matrix and expressing this equation in terms of the rotated coefficient vector, $\mathbf{v}(n)$, we have, for the kth coefficient,

$$E\{v_{n+1}(k)\} = E\{v_n(k)\} - \mu\lambda_k E\{v_{n-1}(k)\} + \mu r_{dx}(k)$$

where λ_k for $k = 1, \ldots, N$ are the eigenvalues of $\mathbf{R}_x$. Since the characteristic equation for $E\{v_n(k)\}$ is

$$1 - z^{-1} + \mu\lambda_k z^{-2} = 0$$

in order for $E\{v_n(k)\}$ to converge in the mean, the roots of the characteristic equation must lie inside the unit circle. Since the roots are at

$$z_k = \tfrac{1}{2}\left\{1 \pm \sqrt{1 - 4\mu\lambda_k}\right\}$$

then the delayed LMS algorithm converges in the mean provided

$$0 < \mu < 1/\lambda_{\max}$$

(b) With $\lambda_k = 1$ and $\mu = 0.1$, the modes of the LMS algorithm behave as

$$(1 - \mu\lambda)^n = (0.9)^n$$

and the time constant is

$$\tau = \frac{1}{\mu\lambda_k} = 10$$

For the delayed LMS, the roots of the characteristic equation are

$$z_k = \tfrac{1}{2}\left\{1 \pm \sqrt{1 - .4}\right\} = 0.887,\ 0.1127$$

Therefore, the slowest decaying mode behaves as $(0.887)^k$ which is approximately the same as the LMS algorithm. Thus, the time constants for delayed LMS with $n_0 = 1$ is about the same as LMS.