EJ.05
Ite 7 17

Advanced Signal Processing
2007

Solutions:

1) a) Bias $B = \theta - E\{\hat{\theta}_N\}$.
If the bias is zero, then the expected value of the estimate is equal to the true value, i.e. $E\{\hat{\theta}_N\} = \theta$ and the estimator is said to be unbiased.

b) An estimator is asymptotically unbiased if an estimate is biased but the bias goes to zero as the number of observations, N goes to infinity, that is

$$\lim_{N\to\infty} E\{\hat{\theta}_N\} = \theta$$

c) An estimate $\hat{\theta}_N$ is said to converge to θ in the mean square sense if

$$\lim_{N\to\infty} E\{|\hat{\theta}_N - \theta|^2\} = 0$$

i) The sample mean estimate is unbiased, since $E\{\hat{m}_x\} = \frac{1}{N}\sum_{n=1}^{N} E\{x[n]\} = m_x$.
Since the variance of the sample mean estimate is

$$var\{\hat{m}_x\} = \frac{1}{N^2}\sum_{n=1}^{N} var\{x[n]\} = \frac{\sigma_x^2}{N}$$

which goes to zero as $N \to \infty$, it follows that the sample mean is a consistent estimator.

d) i) $m_x = E\{x\} = p \cdot 1 + (-1) \cdot (1-p) = 2p - 1$

ii) For an estimator $\hat{m}_x = x[N]$, the mean is

$$E\{\hat{m}_x\} = E\{x[N]\} = 2p - 1$$

and the estimator is unbiased. However, $\hat{m}_x = x[N]$ is not a good estimator of the mean.

iii) The estimate of the mean, $\hat{m}_x$ will either be equal to one, with probability p or it will be equal to minus one, with a probability of $(1-p)$. Therefore the accuracy of the estimate $E\{\hat{m}_x\} = x[N]$ does not improve as the number of observations N increases. The variance of the estimate

$$var\{\hat{m}_x\} = var\{x[N]\} = 4p(1-p)$$

does not decrease with N. The estimator does not converge in the mean square sense and is therefore not consistent.

2) a) The constraints involved in the derivation of the BLUE are the unbiased and linear constraint, that is

$$E\{\hat{\theta}\} = \sum_{n=0}^{N-1} a_n E\{x[n]\} = \theta$$

The vector of parameters $\mathbf{a} = [a_0, \ldots, a_{N-1}]^T$ is found by minimising the variance of the BLUE with the unbiased constraint. Since that is not directly possible, the second constraint is that $E\{x[n]\}$ is linear in θ

$$E\{x[n]\} = s[n]\theta$$

where $s[n]$'s are known.
The variance of the BLUE is

$$var\{\hat{\theta}\} = E\left\{\left(\sum_{n=0}^{N-1} a_n x[n] - E\left\{\sum_{n=0}^{N-1} a_n x[n]\right\}\right)^2\right\} = \mathbf{a}^T\mathbf{C}\mathbf{a}$$

where $\mathbf{C}$ is the correlation matrix.

b) The optimisation problem to find the BLUE is:

- minimise the variance of the BLUE $var\{\hat{\theta}\} = \mathbf{a}^T\mathbf{C}\mathbf{a}$
subject to
- the unbiased constraint $\sum_{n=0}^{N-1} E\{x[n]\} = \theta \Leftrightarrow \sum_{n=0}^{N-1} a_n s[n]\theta = \theta \Leftrightarrow \mathbf{a}^T\mathbf{s} = 1$.

c) To derive the scalar BLUE, we need to minimise the variance $var\{\hat{\theta}\} = \mathbf{a}^T\mathbf{C}\mathbf{a}$ with respect to the constraint $\mathbf{a}^T\mathbf{s} = 1$.
This is a constrained optimisation problem and the Lagrangian function becomes

$$J = \mathbf{a}^T\mathbf{C}\mathbf{a} + \lambda(\mathbf{a}^T\mathbf{s} - 1)$$

The gradient with respect to $\mathbf{a}$ is

$$\frac{\partial J}{\partial \mathbf{a}} = 2\mathbf{C}\mathbf{a} + \lambda\mathbf{s}$$

Setting this equal to zero vector and solving produces

$$\mathbf{a} = -\frac{\lambda}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$$

The Lagrangian multiplier λ is found from the constraint $\mathbf{a}^T\mathbf{s} = 1$, so that the gradient is zero with the constraint satisfied for

$$\mathbf{a}_{opt} = \frac{\mathbf{C}^{-1}\mathbf{s}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}$$

The variance of the BLUE is then $var\{\hat{\theta}\} = \frac{1}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}$.

d) BLUE is applicable to amplitude estimation of known signals in noise.

e) BLUE is inappropriate for problems which are nonlinear in the data, such as for instance estimation of the power of WGN. The MVU estimator in this case is

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n]$$

which is clearly nonlinear in the data.

3) a) The ARMA(p,q) model is given by

$$z[n] = a_1 z[n-1] + a_2 z[n-2] + \cdots + a_p z[n-p] + w[n] + b_1 w[n-1] + \cdots + b_q w[n-q]$$

i) From the equation of an ARMA(p,q) process above, the autocovariance function satisfies the difference equation

$$c_{zz}(k) = a_1 c_{zz}(k-1) + \cdots + a_p c_{zz}(k-p) + c_{zw}(k) + b_1 c_{zw}(k-1) + \cdots + b_q c_{zw}(k-q)$$

Therefore, for $k \geq q+1$, the crosscovariances between z and w vanish and the autovariance becomes

$$c_{zz}(k) = a_1 c_{zz}(k-1) + \cdots + a_p c_{zz}(k-p), \qquad k \geq q+1$$

ii) The ARMA(p,q) process consists of an AR(p) process driven by signal $e[n]$ where $e[n]$ is an $MA(q)$ process. In other words, the ARMA(p,q) process can be considered as an AR(p) process driven by coloured noise $e[n]$. The moving average terms will not affect the denominator of the transfer function and will not influence stationarity of the superimposed autoregressive process.

iii) The power spectrum $P_{zz}(f)$ of an ARMA(p,q) process is given by

$$P_{zz}(f) = \sigma_w^2 \frac{\left|1 + b_1 e^{-\jmath 2\pi f} + \cdots + b_q e^{-\jmath 2\pi q f}\right|^2}{\left|1 - a_1 e^{-\jmath 2\pi f} - \cdots - a_p e^{-\jmath 2\pi p f}\right|^2}$$

b) From the stationarity and invertibility conditions of AR(1) and MA(1) processes, process $z[n]$ is stationary for $-1 < a_1 < 1$ and invertible for $-1 < b_1 < 1$.

i) and ii) From question a) part i) we know that the autocorrelation coefficients of an ARMA(p,q) process are equal to the autocorrelation coefficients of an AR(p) process for $k \geq q+1$. From the autocovariance function, dividing by $c(0)$, for an ARMA(1,1) process we have

$$\begin{aligned} \rho_0 &= 1 \\ \rho_1 &= \frac{(1 + a_1 b_1)(a_1 + b_1)}{1 + b_1^2 + 2 a_1 b_1} \\ \rho_k &= a_1 \rho_{k-1}, \qquad k \geq 2 \end{aligned}$$

4) a) Solving the linear mean square estimation problem starts with differentiating J with respect to a and b and setting the derivatives to zero, as

$$\frac{\partial J}{\partial a} = -2E\{(y - ax - b)x\} = -2E\{xy\} + 2aE\{x^2\} + 2bm_x = 0$$

$$\frac{\partial J}{\partial b} = -2E\{y - ax - b\} = -2m_y + 2am_x + 2b = 0$$

It then follows that

$$a = \frac{E\{xy\} - m_x m_y}{\sigma_x^2}$$

$$b = \frac{E\{x^2\}m_y - E\{xy\}m_x}{\sigma_x^2}$$

Since

$$E\{xy\} = a\sigma_x^2 + m_x m_y$$

we have

$$a = \rho_{xy}\frac{\sigma_y}{\sigma_x}$$

$$b = m_y - am_x$$

b) Using the values for a and b from part a), the minimum mean square error can be evaluated as

$$E\left\{(y - \hat{y})^2\right\} = \sigma_y^2 - a^2\sigma_x^2 = \sigma_y^2(1 - \rho_{xy}^2)$$

Notice that if x and y are uncorrelated, then $a = 0$ and $b = E\{y\}$.

c) The advantages of using a linear mean square estimator are:
- The parameters a and b depend only on the second order moments of x and y and not on the joint density functions.
- The equations for solving for a and b are linear, hence the computational complexity is low.
- For Gaussian random variables, the optimum mean square estimate is linear.

d) The linear mean square estimate assumes fixed coefficients a and b and the problem is solved in a block fashion, that is, taking into account simultaneously all the available data.
Linear FIR filter, trained by the LMS algorithm, on the other hand deals only with a portion of data which is present in its tap inputs, the estimation is recursive, and the coefficients are adaptive.

5) a)

$$J = \frac{1}{2}E\{e^2\} = \frac{1}{2}E\{d^2\} - E\{\sum_{k=1}^{N} w_k x_k d\} + \frac{1}{2}E\{\sum_{j=1}^{N}\sum_{k=1}^{N} w_j w_k x_j x_k\}$$

Therefore

$$J = \frac{1}{2}r_d - \sum_{k=1}^{N} w_k r_{dx}(k) + \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} w_j w_k r_{xx}(j,k)$$

Differentiating J with respect to w_k and setting to zero we obtain

$$\sum_{j=1}^{N} w_{oj} r_{xx}(j,k) = r_{xd}(k), \qquad k = 1, 2, \ldots, N$$

where w_{oj} are the optimal weights.

b) To avoid the matrix inversion involved in the Wiener filter, we can make the weights time varying and adapt them in an iterative fashion. In that case

$$\Delta w_k(n) = -\eta \nabla_{w_k} J(n), \qquad k = 1, \ldots, N$$

and

$$w_k(n+1) = w_k(n) + \Delta w_k(n)$$

Therefore

$$w_k(n+1) + w_k(n) + \eta \left[r_{dx} - \sum_{j=1}^{N} w_j(n) r_{xx}(j,k) \right]$$

The method of steepest descent is exact in the sense that there are no approximations made in its derivation.

c) If we use the instantaneous estimates instead of the exact estimates of the autocorrelation and crosscorelation functions, we have

$$r_{xx}(j,k;n) = x_j(n)x_k(n)$$
$$r_{dx}(k;n) = x_k(n)d(n)$$

and the LMS algorithm becomes

$$w_k(n) = w_k(n-1) + \eta e(k) x_k(n), \qquad k = 1, \ldots, N$$

i) The learning rate defines the step towards the global minimum on the error surface. The convergence in the mean is attained provided that

$$0 < \eta < \frac{2}{\lambda_{max}}$$

where λ_{max} is the largest eigenvalue of the autocorrelation matrix. Convergence in the mean square is preserved if the learning rate is positive and its value is less than the twice the inverse of the total input power.

ii) The LMS is computationally much simpler, can deal with both the stationary and nonstationary data and is very robust. However, due to the instantaneous estimates of the second order quantities from the Wiener filter, its steady state error can be considerable and the convergence is relatively slow. The LMS is a suboptimal iterative solution to the Wiener filtering problem.

iii) A multidimensional plot of the cost function J versus the weights constitutes the error–performance surface, or simply the error surface of the filters. It is bowl–shaped with a well defined bottom or global minimum point. The error surface of NLMS has its contours which are ideally concentric circles, which is not the case with the error surface of LMS, where the contours are elliptic.