Spectral Estimation and Adaptive Signal Processing

Problem and Answer Sets

For generality, some of the solutions have been derived for complex-valued data.

To obtain the corresponding solution for real-valued data, simply replace the Hermitian transpose operator $(\cdot)^H$ with the vector/matrix transpose operator $(\cdot)^T$.

Problem Set I : Background

1.1 The DFT of a sequence $x(n)$ of length N may be expressed in a matrix form as follows

$$\mathbf{X} = \mathbf{W}\mathbf{x}$$

where $\mathbf{x} = \left[\begin{array}{c} x(0), x(1), \ldots, x(N-1) \end{array} \right]^T$ is a vector containing the signal values and $\mathbf{X}$ is a vector containing the DFT coefficients $X(k)$.

(a) Find the matrix $\mathbf{W}$.

(b) What are the properties of $\mathbf{W}$?

(c) What is the inverse of $\mathbf{W}$?

Solution

(a) The DFT sequence of length N is given by

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}nk} = \sum_{n=0}^{N-1} x(n) W_N^{nk}$$

where $W_N = e^{-j\frac{2\pi}{N}}$. If we define

$$\mathbf{w}_k^H = \left[\begin{array}{c} 1, W_N^k, W_N^{2k}, \ldots, W_N^{k(N-1)} \end{array} \right]$$

where $X(k)$ is the inner product

$$X(k) = \mathbf{w}_k^H \cdot \mathbf{x} \tag{1}$$

Arranging the DFT coefficients in a vector we have

$$\mathbf{X} = \begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_0^H \mathbf{x} \\ \mathbf{w}_1^H \mathbf{x} \\ \vdots \\ \mathbf{w}_{N-1}^H \mathbf{x} \end{bmatrix} = \mathbf{W}\mathbf{x}$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_0^H \\ \mathbf{w}_1^H \\ \vdots \\ \mathbf{w}_{N-1}^H \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \ldots & 1 \\ 1 & W_N & W_N^2 \ldots & W_N^{N-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & W_N^{N-1} & W_N^{2(N-1)} \ldots & W_N^{(N-1)^2} \end{bmatrix}$$

(b) The matrix $\mathbf{W}$ is symmetric and nonsingular. In addition, due to orthogonality of the complex exponentials,

$$\mathbf{w}_k^H \cdot \mathbf{w}_l = \sum_{n=0}^{N-1} e^{-j\frac{2\pi}{N}(k-l)} = \begin{cases} N & \text{if } k = l, \\ 0 & \text{if } k \neq l. \end{cases} \tag{2}$$

it follows that $\mathbf{W}$ is orthogonal.

(c) Due to orthogonality of $\mathbf{W}$, the inverse is

$$\mathbf{W}^{-1} = \frac{1}{N}\mathbf{W}^H \tag{3}$$

This explains the form for the inverse Discrete Fourier Transform (DFT)

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}kn}$$

1.2 Consider a problem of trying to model a sequence $x(n)$ as a sum of constant plus a complex exponential of frequency w_0, that is

$$\hat{x}(n) = c + ae^{jnw_0}, \qquad n = 0, 1, \ldots, N-1$$

where a and c are unknowns. We may express the problem of finding the values of c and a as one of solving a set of overdetermined linear equations

$$\begin{bmatrix} 1 & 1 \\ 1 & e^{jw_0} \\ \vdots & \vdots \\ 1 & e^{j(N-1)w_0} \end{bmatrix} \begin{bmatrix} c \\ a \end{bmatrix} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}$$

(a) Find the least squares solution for a and c.

(b) If N is even and $w_0 = \frac{2\pi k}{N}$ for some integer k, find the least squares solution for c and a.

Solution

(a) Assuming that $w_0 \neq 0, 2\pi, \ldots$, the columns of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & e^{jw_0} \\ \vdots & \vdots \\ 1 & e^{j(N-1)w_0} \end{bmatrix}$$

are linearly independent, and the leasta squares (LS) solution for a and c (standard LS) is given by

$$\begin{bmatrix} c \\ a \end{bmatrix} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{x}$$

since

$$\mathbf{A}^H\mathbf{A} = \begin{bmatrix} N & \sum_{n=0}^{N-1} e^{jnw_0} \\ \sum_{n=0}^{N-1} e^{-jnw_0} & N \end{bmatrix} = \begin{bmatrix} N & \frac{1-e^{jNw_0}}{1-e^{jw_0}} \\ \frac{1-e^{-jNw_0}}{1-e^{-jw_0}} & N \end{bmatrix}$$

Therefore, the inverse of $(\mathbf{A}^H\mathbf{A})$ is

$$(\mathbf{A}^H\mathbf{A})^{-1} = \frac{1}{N^2 - \frac{1-\cos Nw_0}{1-\cos w_0}} \begin{bmatrix} N & -\frac{1-e^{jNw_0}}{1-e^{jw_0}} \\ -\frac{1-e^{-jNw_0}}{1-e^{-jw_0}} & N \end{bmatrix}$$

and we have

$$\begin{bmatrix} c \\ a \end{bmatrix} = \frac{1}{N^2 - \frac{1-\cos Nw_0}{1-\cos w_0}} \begin{bmatrix} N & -\frac{1-e^{jNw_0}}{1-e^{jw_0}} \\ -\frac{1-e^{-jNw_0}}{1-e^{-jw_0}} & N \end{bmatrix} \begin{bmatrix} \sum_{n=0}^{N-1} x(n) \\ \sum_{n=0}^{N-1} x(n)e^{-jw_0 n} \end{bmatrix}$$

which becomes

$$\begin{bmatrix} c \\ a \end{bmatrix} = \frac{1}{N^2 - \frac{1-\cos Nw_0}{1-\cos w_0}} \begin{bmatrix} N\sum_{n=0}^{N-1} x(n) - \frac{1-e^{jNw_0}}{1-e^{jw_0}} \sum_{n=0}^{N-1} x(n)e^{-jnw_0} \\ N\sum_{n=0}^{N-1} x(n)e^{-jnw_0} - \frac{1-e^{-jNw_0}}{1-e^{-jw_0}} \sum_{n=0}^{N-1} x(n) \end{bmatrix}$$

(b) If $w_0 = \frac{2\pi k}{N}$ and $k \neq 0$ then

$$\frac{1-e^{jNw_0}}{1-e^{jw_0}} = \frac{1-e^{-jNw_0}}{1-e^{-jw_0}} = 0$$

and

$$\frac{1-\cos Nw_0}{1-\cos w_0} = 0$$

Therefore, we have

$$\begin{bmatrix} c \\ a \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{n=0}^{N-1} x(n) \\ \frac{1}{N}\sum_{n=0}^{N-1} x(n)e^{-jw_0 n} \end{bmatrix}$$

1.3 Let $\mathbf{A} > 0$ and $\mathbf{B} > 0$ be positive definite matrices. Prove or disprove the following statements.

(a) $\mathbf{A}^2 > 0$.

(b) $\mathbf{A}^{-1} > 0$.

(c) $\mathbf{A} + \mathbf{B} > 0$.

Solution

(a) Let $\mathbf{v}_k$ be an eigenvector and λ_k the corresponding eigenvalue of $\mathbf{A}$. Since

$$\mathbf{A}^2\mathbf{v}_k = \mathbf{A}(\mathbf{A}\mathbf{v}_k) = \lambda_k \mathbf{A}\mathbf{v}_k = \lambda_k^2 \mathbf{v}_k$$

then $\mathbf{v}_k$ is an eigenvector of $\mathbf{A}^2$ and λ_k^2 is the corresponding eigenvalue. If $\mathbf{A} > 0$, then $\lambda_k > 0$. Therefore, $\lambda_k^2 > 0$ and it follows that $\mathbf{A}^2 > 0$.

(b) If $\mathbf{A} > 0$, then eigenvalues of $\mathbf{A}$ are positive, $\lambda_k > 0$; $\mathbf{A}^{-1}$ also exists and eigenvalues of $\mathbf{A}^{-1}$ are λ_k^{-1}. Since $\lambda_k > 0$, therefore $\lambda_k^{-1} > 0$ and therefore $\mathbf{A}^{-1} > 0$

(c) Let $\mathbf{v} \neq 0$ be an arbitrary (complex) vector. Then

$$\mathbf{v}^H(\mathbf{A} + \mathbf{B})\mathbf{v} = \mathbf{v}^H\mathbf{A}\mathbf{v} + \mathbf{v}^H\mathbf{B}\mathbf{v}$$

If $\mathbf{A} > 0$ and $\mathbf{B} > 0$, then

$$\mathbf{v}^H\mathbf{A}\mathbf{v} > 0;\ \mathbf{v}^H\mathbf{B}\mathbf{v} > 0$$

Therefore,

$$\mathbf{v}^H(\mathbf{A} + \mathbf{B})\mathbf{v} > 0$$

and it follows that $(\mathbf{A} + \mathbf{B}) > 0$.

1.4 (a) Find the eigenvalues and eigenvectors of a real 2x2 symmetric Toeplitz matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$$

(b) Find the eigenvalues and eigenvectors of a complex 2x2 Hermitian matrix

$$\mathbf{A} = \begin{bmatrix} a & b^* \\ b & a \end{bmatrix}$$

Solution

(a) The eigenvalues of roots of characteristic equation

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (a - \lambda)^2 - b^2 = 0$$

Expanding the quadratic in λ we have

$$\lambda^2 - 2a\lambda + (a^2 - b^2) = 0$$

which gives $\lambda_1 = a + b$ and $\lambda_2 = a - b$. The eigenvectors can be found by

$$\mathbf{A}\mathbf{v}_k = \lambda_k\mathbf{v}_k$$

For the first eigenvector $\mathbf{v}_1$, we have

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = (a+b) \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix}$$

which gives $v_{11} = v_{12}$, or

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Similarly, eigenvector $\mathbf{v}_2$ can be found to be

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

(b) With

$$\mathbf{A} = \begin{bmatrix} a & b^* \\ b & a \end{bmatrix}$$

the eigenvalues are the roots of the equation

$$\det(\mathbf{A} - \lambda \mathbf{I}) = (a - \lambda)^2 - |b|^2 = 0$$

or

$$\lambda^2 - 2a\lambda + (a^2 - |b|^2) = 0$$

or,

$$[\lambda - (a + |b|)][\lambda - (a - |b|)] = 0$$

Thus, $\lambda_1 = a + |b|$ and $\lambda_2 = a - |b|$

1.5 a 2x2 matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

(a) Find the eigenvalues and eigenvectors of $\mathbf{A}$.

(b) Are the eigenvectors unique? Are they linearly independent? Are they orthogonal?

(c) Diagonalize $\mathbf{A}$ i.e. find $\mathbf{V}$ and $\mathbf{D}$ such that

$$\mathbf{V}^H \mathbf{A} \mathbf{V} = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix.

Solution

(a) The eigenvalues are the roots of the characteristic equation

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \lambda^2 + 1 = 0$$

where $\lambda = \pm j$. The eigenvectors corresponding to $\lambda_1 = j$ satisfy the equation

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = j \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

which implies that $v_2 = jv_1$. The normalized eigenvector is

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ j \end{bmatrix}$$

Similarly, the eigenvector corresponding to $\lambda_2 = -j$ is

$$\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -j \end{bmatrix}$$

(b) The eigenvectors are unique, linearly independent and orthogonal, that is

$$< \mathbf{v}_1, \mathbf{v}_2 >= \mathbf{v}_1^H \mathbf{v}_2 = 0$$

(c) With $\mathbf{V}$ the matrix of normalized eigenvectors,

$$\mathbf{V} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}$$

we have

$$\mathbf{V}^H \mathbf{A} \mathbf{V} = \mathbf{D}$$

where

$$\mathbf{D} = \begin{bmatrix} j & 0 \\ 0 & -j \end{bmatrix}$$

1.6 Find the matrix whose eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 4$ with eigenvectors $\mathbf{v}_1 = [3, 1]^T$ and $\mathbf{v}_2 = [2, 1]^T$

Solution

For the given information, we have

$$\mathbf{A} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} ; \; \mathbf{A} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

Let

$$\mathbf{A} = [\mathbf{a}_1, \; \mathbf{a}_2]$$

Then we have

$$3\mathbf{a}_1 + \mathbf{a}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

and

$$2\mathbf{a}_1 + \mathbf{a}_2 = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

Subtracting these two equations gives

$$\mathbf{a}_1 = \begin{bmatrix} -5 \\ -3 \end{bmatrix}$$

Also, we have

$$\mathbf{a}_2 = \begin{bmatrix} 8 \\ 4 \end{bmatrix} - 2\mathbf{a}_1 = \begin{bmatrix} 18 \\ 10 \end{bmatrix}$$

Therefore,

$$\mathbf{A} = \begin{bmatrix} -5 & 18 \\ -3 & 10 \end{bmatrix}$$

1.7 Check if the following matrices are positive (semi)-definite:

a) $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

b) $\mathbf{B} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

c) $\mathbf{C} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

Solution

We need to check whether $\mathbf{a}^T\mathbf{R}\mathbf{a} \geq 0$, for any $\mathbf{a} \in \mathbb{R}^N \times 1$ and $\mathbf{R}$ of size $N \times N$.

a) For a vector $\mathbf{a} = [a_1, a_2]^T$, we have

$$[a_1,\ a_2]\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad = \quad [a_1,\ a_2]\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = a_1^2 + a_2^2$$

This is positive, and thus $\mathbf{A}$ is positive definite.

b) For any non-zero vector $\mathbf{x} = [x_1,\ x_2,\ x_3]^T$, we have

$$\begin{aligned} \mathbf{x}^T\mathbf{B}\mathbf{x} &= [x_1,\ x_2,\ x_3]\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 - x_2, & -x_1 + 2x_2 - x_3, & -x_2 + 2x_3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2 \end{aligned}$$

This is a sum of squares and hence non-negative, and therefore $\mathbf{B}$ is positive semidefinite (=0 for $x_1, x_2, x_3 = 0$).

c) Using the same approach as above, the quadratic form $\mathbf{b}^T\mathbf{C}\mathbf{b} = -2 < 0$, and $\mathbf{C}$ is not positive definite.

Problem Set II : Estimation Theory - Background

2.1 Given $x[0], x[1], ..., x[N-1]$ which are independent and identically distributed (IID) according to the $\mathcal{N}(\mu_x, \sigma_x^2)$ distribution, it is desired to estimate $\hat{\mu}_x$ by the sample mean

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

(a) Prove that $\hat{\mu}_x$ is an unbiased estimator.

(b) Find the variance of the estimator.

(c) Show that the Cramer-Rao lower bound (CRLB) is given by

$$var(\mu_x) \geqslant \sigma_x^2/N$$

Comment on the efficiency of sample mean estimator for such data.

Solution

(a) Mean:-

$$\mathrm{E}\{\hat{\mu}_x\} = E\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} = \frac{1}{N}\sum_{n=0}^{N-1} E\{x[n]\} = \frac{1}{N}\sum_{n=0}^{N-1}\mu_x = \mu_x$$

(b) Variance:-

$$\begin{aligned}
\mathrm{var}\big\{\hat{\mu}_x\big\} &= E\big\{(\hat{\mu}_x - \mathrm{E}\{\hat{\mu}_x\})^2\big\} = E\Big\{\big(\frac{1}{N}\sum_{n=0}^{N-1} x[n] - \mu_x\big)^2\Big\} \\
&= \frac{1}{N^2}\mathrm{E}\Big\{\sum_{n=0}^{N-1}(x[n]-\mu_x)^2 + \sum_{n=0}^{N-1}\sum_{m=0}^{N-1}(x[n]-\mu_x)(x[m]-\mu_x)\Big\} \\
&= \frac{1}{N^2}\Big(N\sigma_x^2 + \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} E\underbrace{\big\{(x[n]-\mu_x)(x[m]-\mu_x)\big\}}_{=0\ \ \text{(independence assumption)}}\Big) = \sigma_x^2/N
\end{aligned}$$

(c) CRLB:-

$$\mathrm{var}\{\hat{\mu}_x\} \geq \frac{1}{E\{\big(\frac{\partial}{\partial\theta}\ln P(\mathbf{x},\theta)\big)^2\}}$$

$$\begin{aligned}
P(\mathbf{x};\mu_x) &= \prod_{i=0}^{N-1}\frac{1}{\sqrt{2\pi}\sigma_x}exp\Big[-1/2\Big(\frac{x[i]-\mu_x}{\sigma_x}\Big)^2\Big] \\
\ln(P(\mathbf{x};\mu_x)) &= \ln(\frac{1}{2\pi^{N/2}\sigma_x^N}) - \frac{1}{2\sigma_x^2}\sum_{i=0}^{N-1}(x[i]-\mu_x)^2 \\
\frac{\partial}{\partial\mu_x}\ln P(\mathbf{x};\mu_x) &= \frac{-1}{2\sigma_x^2} - 2\sum_{i=0}^{N-1}(x[i]-\mu_x) = \frac{1}{\sigma_x^2}\sum_{i=0}^{N-1}(x[i]-\mu_x) \\
\Rightarrow &\quad E\Big\{\frac{\partial}{\partial\mu_x}\ln\big(P(\mathbf{x};\mu_x)\big)^2\Big\} = \frac{1}{\sigma_x^4}\mathrm{E}\Big\{\big[\sum_{i=\phi}^{N-1}(x[i]-\mu_x)\big]^2\Big\} = \frac{N}{\sigma_x^2} \\
\Rightarrow &\quad \text{CR Lower Bound Variance is} \quad Var(\hat{\mu}_x) \geq \frac{\sigma_x^2}{N}
\end{aligned}$$

This sample mean estimate is efficient for IID processes.

2.2 Let x be a random variable with mean m_x and variance σ_x^2. Let x_i for $i = 1, 2, \ldots, N$ be N independent measurements of the random variable x.

(a) With $\hat{m}_x$ the sample mean defined by

$$\hat{m}_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

determine whether or not the sample variance

$$\hat{\sigma}_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{m}_x)^2$$

is unbiased, i.e., is $E\{\hat{\sigma}_x^2\} = \sigma_x^2$?

(b) If x is a Gaussian random variable, find the variance of the sample variance, $E\{(\hat{\sigma}_x^2 - E\{\hat{\sigma}_x^2\})^2\}$.

Solution

(a) The expected value of the sample variance is

$$E\{\hat{\sigma}_x^2\} = E\left\{\frac{1}{N}\sum_{i=1}^{N}(x_i - \frac{1}{N}\sum_{j=1}^{N}x_j)^2\right\} = \frac{1}{N}\sum_{i=1}^{N}E\left\{\left[(x_i - m_x) - \frac{1}{N}\sum_{j=1}^{N}(x_j - m_x)\right]^2\right\}$$

Expanding the square we have

$$E\{\hat{\sigma}_x^2\} = \frac{1}{N}\sum_{i=1}^{N}E\left\{(x_i - m_x)^2 - \frac{2}{N}\sum_{j=1}^{N}(x_i - m_x)(x_j - m_x) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(x_i - m_x)(x_j - m_x)\right\}$$

Since the measurements are assumed to be independent, then

$$E\{(x_i - m_x)(x_j - m_x)\} = \begin{cases} \sigma_x^2 & ; \ \text{i=j} \\ 0 & ; \ \text{i}\neq\text{j} \end{cases}$$

and the expression for $\hat{\sigma}_x^2$ becomes

$$E\{\hat{\sigma}_x^2\} = \frac{1}{N}\sum_{i=1}^{N}\{\sigma_x^2 - \frac{2}{N}\sigma_x^2 + \frac{1}{N^2}N\sigma_x^2\} = \sigma_x^2(1 - \frac{1}{N}) = \sigma_x^2\frac{N-1}{N}$$

Therefore, although the sample variance is biased, it is asymptotically unbiased.

(b) Finding the variance of the sample variance directly is very tedious. A simpler way is as follows. With

$$\hat{\sigma}_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{m}_x)^2$$

it is well-known that

$$\frac{N\hat{\sigma}_x^2}{\sigma_x^2} = \sum_{i=1}^{N}\left(\frac{x_i - \hat{m}_x}{\sigma_x}\right)^2$$

is a Chi-square random variable with $(n-1)$ degrees of freedom, which has a variance of $2(n-1)$. Therefore,

$$Var\left(\frac{N\hat{\sigma}_x^2}{\sigma_x^2}\right) = 2(N-1)$$

and, consequently, we have

$$Var(\hat{\sigma}_x^2) = 2\frac{\sigma_x^4}{N^2}(N-1)$$

2.3 It is desired to predict the real WSS random process $x[n]$ based on the previous sample $x[n-1]$ by using the linear predictor

$$\hat{x}[n] = -\alpha_1 x[n-1]$$

The value of parameter α_1 is chosen to minimize the MSE or prediction error power

$$MSE = E\{(x[n] - \hat{x}[n])^2\}$$

(a) Employ the orthogonality principle to find the optimal prediction parameter α_1 and the minimum prediction error power.

Solution

Orthogonality Principle states that the estimation error is orthogonal to the data, that is $(x[n] - \hat{x}[n]) \perp x[n-1]$, that is

$$\begin{aligned} \Rightarrow E\{(x[n] - \hat{x}[n])x[n-1]\} = 0 \\ E\{(x[n] + \alpha_1 x[n-1])x[n-1]\} = 0 \\ \Rightarrow r_{xx}(1) = -\alpha_1 r_{xx}(0) \\ \Rightarrow \alpha_1 = -r_{xx}(1)/r_{xx}(0) \end{aligned}$$

Since for the AR(1) model we have

$$x[n] = \alpha_1 x[n-1] + w[n], \qquad w[n] \sim \mathcal{N}(0, \sigma_w^2)$$

from the Yule-Walker normal equation we have

$$\sigma_w^2 \quad = \quad r_{xx}[0] + \alpha_1 r_{xx}[1] = r_{xx}[0]\Big(1 - \Big[\frac{r_{xx}[1]}{r_{xx}[0]}\Big]^2\Big)$$

2.4 Consider the random process

$$x(n) = A\cos(n\omega + \phi) + w(n)$$

where $w(n)$ is zero mean white Gaussian noise with a variance σ_w^2. For each of the following cases, find the autocorrelation sequence and if the process is WSS, find the power spectrum.

(a) A is a Gaussian random variable with zero mean and variance σ_A^2 and both ω and ϕ are constants.

(b) ϕ is uniformly distributed over the interval $[-\pi, \pi]$ and both A and ω are constants.

(c) ω is a random variable that is uniformly distributed over the interval $[\omega_0 - \Delta, \omega_0 + \Delta]$ and both A and ϕ are constants.

Solution

(a) When ω and ϕ are constants, then

$$r_x(k,l) = E\{x(k)x(l)\} = E\{A\cos(k\omega + \phi)A\cos(l\omega + \phi)\} + \sigma_w^2 \delta(k-l)$$

Thus,

$$\begin{aligned} r_x(k,l) &= E\{A^2\}\cos(k\omega + \phi)\cos(l\omega + \phi) + \sigma_w^2\delta(k-l) \\ &= \sigma_A^2 \cos(k\omega + \phi)\cos(l\omega + \phi) + \sigma_w^2\delta(k-l) \end{aligned}$$

Note that since $r_x(k,l)$ does not depend on the difference $(k-l)$, then $x(n)$ is not wide-sense stationary, and the power spectrum is not defined for this process.

(b) When A and ω are constants and ϕ is a random variable that is uniformly distributed over the interval $[-\pi, \pi]$, then the autocorrelation is,

$$\begin{aligned} r_x(k,l) &= E\{A^2\cos(k\omega+\phi)\cos(l\omega+\phi)\} + \sigma_w^2\delta(k-l) \\ &= \frac{1}{2}A^2E\{\cos[(k+l)\omega+2\phi]\} + \frac{1}{2}A^2E\{\cos(k-l)\omega\} + \sigma_w^2\delta(k-l) \end{aligned}$$

However, since $E\{\cos[(k+l)\omega+2\phi]\} = 0$, then the autocorrelation is

$$r_x(k,l) = \frac{1}{2}A^2\cos(k-l)\omega + \sigma_w^2\delta(k-l)$$

Therefore, $r_x(k,l)$ depends on the difference $(k-l)$, and the process is WSS. The power spectrum is

$$P_x(e^{j\omega}) = \frac{\pi A^2}{2}\delta(\omega-\omega) + \frac{\pi A^2}{2}\delta(\omega+\omega) + \sigma_w^2$$

(c) As in parts (a) and (b), the autocorrelation of the process $x(n)$ is

$$r_x(k,l) = E\{x(k)x(l)\} = E\{A\cos(k\omega+\phi)A\cos(l\omega+\phi)\} + \sigma_w^2\delta(k-l)$$

In this case, however, ω is a random variable, and the expectation of the product of the cosines is

$$E\{A\cos(k\omega+\phi)A\cos(l\omega+\phi)\} = A^2E\left\{\frac{1}{2}\cos[(k-l)\omega] + \frac{1}{2}\cos[(k+l)\omega+2\phi]\right\}$$

Since ω is uniformly distributed over the interval $[\omega_0-\Delta, \omega_0+\Delta]$, the expectation of the first term is

$$\begin{aligned} E\{\cos[(k-l)\omega]\} &= \frac{1}{2\Delta}\int_{\omega-\Delta}^{\omega+\Delta}\cos[(k-l)\omega]d\omega \\ &= \frac{1}{2\Delta(k-l)}\left[\sin[(k-l)(\omega_0+\Delta)] - \sin[(k-l)(\omega_0-\Delta)]\right] \end{aligned}$$

With ϕ a constant, the expectation of the second term is

$$E\{\cos[(k+l)\omega+2\phi]\} = \frac{1}{2\Delta(k+l)}\left[\sin[(k+l)(\omega_0+\Delta+2\phi)] - \sin[(k+l)(\omega_0-\Delta+2\phi)]\right]$$

Therefore, $x(n)$ is not WSS. However, if ϕ is a random variable that is uniformly distributed over the interval $[-\pi, \pi]$, then this second expectation is zero, and the autocorrelation becomes

$$\begin{aligned} r_x(k,l) &= \frac{A^2}{4\Delta(k-l)}\left\{\sin[(k-l)(\omega_0+\Delta)] - \sin[(k-l)(\omega_0-\Delta)]\right\} \\ &= \frac{A^2}{2\Delta(k-l)}\sin[(k-l)\Delta]\cos[(k-l)\omega_0] \end{aligned}$$

and the process is WSS. With an autocorrelation sequence given by

$$r_x(k) = \frac{\pi A^2}{2\Delta}\frac{\sin k\Delta}{\pi k}\cos k\omega_0$$

using the DTFT pair

$$r_x(k) = \frac{\sin k\Delta}{\pi k} \Leftrightarrow P_x(e^{j\omega}) = \begin{cases} 1 & ; \quad |\omega| \le \Delta \\ 0 & ; \quad else \end{cases}$$

it follows that the power spectrum of $x(n)$ is

$$P_x(e^{j\omega}) = \begin{cases} \frac{\pi A^2}{4\Delta} & ; \quad \omega_0 - \Delta \leq |\omega| \leq \omega_0 + \Delta \\ 0 & ; \quad else \end{cases}$$

2.5 For a real sinusoidal process from the previous example:

(a) Find the temporal autocorrelation function (ACF)

$$\hat{r}_{xx}[n] = \frac{1}{2M+1} \sum_{n=-M}^{M} x[n]x[n+k]$$

as $M \to \infty$. Is the random process autocorrelation ergodic?

(b) Verify the identity

$$\sum_{m=0}^{N-1} exp(-j2\pi f_0 m) = exp(-j\pi f_0[N-1]) \frac{\sin \pi f_0 N}{\sin \pi f_0}$$

Solution

(a)

$$\begin{aligned} r_{xx}[k] &= \frac{1}{2M+1} \sum_{n=-M}^{M} A^2 \cos(2\pi f_0 n + \phi) \cos(2\pi f_0[n+k] + \phi) \quad \text{for } M \to \infty \\ &= \frac{1}{2M+1} \frac{A^2}{2} \sum_{n=-M}^{M} \cos(2\pi f_0[2n+k] + 2\phi) + \cos(2\pi f_0 k) \quad \text{for } M \to \infty \\ &= A^2 \cos(2\pi f_0 k) \end{aligned}$$

Therefore, the process is Ergodic.

(b)

$$\begin{aligned} \sum_{m=0}^{N-1} exp(-j2\pi f_0 m) &= \frac{(1 - exp(-j2\pi f_0 N))}{(1 - exp(-j2\pi f_0))} = \frac{exp(-j\pi f_0 N)(exp(j\pi f_0 N) - exp(-j\pi f_0 N))}{exp(-j\pi f_0)(exp(j\pi f_0) - exp(-j\pi f_0))} \\ &= exp(-j\pi f_0[N-1]) \frac{\sin(\pi f_0 N)}{\sin(\pi f_0)} \end{aligned}$$

2.6 Least Squares Spectral Approximation. Assume we are given an ACS $\{r(k)\}$ or, equivalently, the PSD function $\phi(\omega)$. We wish to find a finite-impulse response (FIR) filter $H(\omega) = h_0 + h_1 e^{-j\omega} + \cdots + h_m e^{-jm\omega}$, whose input $x(n)$ is zero-mean unit-variance white noise and such that the output sequence $y(n)$ has PSD $\phi_y(\omega)$ "close" to $\phi(\omega)$. Specifically, we wish to find $\mathbf{h} = [h_0, \ldots, h_m]$ so that the approximation error

$$\epsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} [\phi(\omega) - \phi_y(\omega)]^2 d\omega \tag{4}$$

is minimum.

- Show that ϵ is a quartic (fourth-order) function of $\mathbf{h}$ and that thus no simple closed-form solution $\mathbf{h}$ exists to minimise (**??**).
- We attempt to reparametrise the minimisation problem as follows: We note that $r_y(k) = 0$ for $|k| > m$; thus,

$$\phi_y(\omega) = \sum_{k=-m}^{m} r_y(k) e^{j\omega k}. \tag{5}$$

Equation (**??**), and the fact that $r_y(-k) = r_y^*(k)$, mean that $\phi(\omega)$ is a function of $\mathbf{g} = [r_y(0), \ldots, r_y(m)]^T$.

Show that the minimisation problem in (**??**) is quadratic in $\mathbf{g}$; it thus admits a closed-form solution. Show that the vector $\mathbf{g}$ that minimises ϵ in Equation (**??**) gives

$$r_y(k) = \begin{cases} r_y(k) & \text{if } |k| \leq m, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

- Can you identify any problems with the "solution" (**??**)?

Solution

- By Parseval's equality

$$\epsilon = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\phi(\omega) - \phi_y(\omega)]^2 d\omega = \sum_{k=-\infty}^{\infty} |r(k) - r_y(k)|^2, \tag{7}$$

where

$$r_y(k) = E\{y(n)y^*(n-k)\} \quad = \quad E\left\{\sum_{u=0}^{m} h_u x(n-u) \sum_{v=0}^{m} h_v^* x^*(n-v)\right\} \tag{8}$$

$$= \begin{cases} \sigma^2 \sum_{u=k}^{m} h_u h_{u-k}^* & 0 \leq k \leq m, \\ \sigma^2 \sum_{u=-k}^{m} h_{u+k} h_u^* & -m \leq k \leq -1. \end{cases} \tag{9}$$

Since $r_y(k)$ is quadratic in $\{h_k\}$, ϵ, which s quadratic in $\{r_y(k)\}$, is quartic in $\{h_k\}$.

- If $\phi_y(\omega) = \sum_{k=-m}^{m} r_y(k)e^{j\omega k}$, then

$$\epsilon = \sum_{k=-\infty}^{\infty} |r(k) - r_y(k)|^2 = \sum_{|k|>m} |r(k)|^2 + \sum_{|k|\leq m} |r(k) - r_y(k)|^2, \tag{10}$$

which is quadratic in $\mathbf{g}$. We see that ϵ is minimised by choosing $r(k) = r_y(k)$ for $|k| \leq m$:

$$\mathbf{g} = [r_y(0), \ldots, r_y(m)]^T. \tag{11}$$

- The problem is that $\phi_y(\omega)$ may not be greater than zero for all ω, since it is associated with a truncated ACS sequence.

2.7 Unbiased ACS Estimates Can Lead to Negative Spectral Estimates. The use of unbiased ACS estimates of the form

$$\hat{r}(k) = \frac{1}{N-k}\sum_{n=k}^{N-1} y(n)y^*(n-k), \quad 0 \leq k \leq N-1 \tag{12}$$

in the correologram spectral estimate

$$\hat{\phi}_c(\omega) = \sum_{k=-(N-1)}^{(N-1)} \hat{r}(k)e^{-j\omega k}, \tag{13}$$

then negative spectral estimates could result. Find an example data sequence $\{y(n)\}_{n=1}^{N}$ that gives such a negative spectral estimate.

Solution

A simple example is $y(n) = \{1, 1.1, 1\}$, whose unbiased ACS estimate is $\hat{r}(k) = \{1-07, 1-1, 1\}$ for $k = 0, 1, 2$. Since $|\hat{r}(1)| > |\hat{r}(0)|$, the sequence cannot be positive semi-definite, and $\hat{\phi}(\omega)$ is not greater that zero for all frequencies. In fact,

$$\hat{\phi}(\omega) = 1.07 + 2.2\cos\omega + 2\cos 2\omega, \tag{14}$$

which is negative for example at $\omega = 0.6\pi$.

2.8 Verify the inverse of real symmetric Toeplitz matrix

$$\mathbf{A} = \begin{pmatrix} 1 & -\alpha & \alpha^2 \\ -\alpha & 1 & -\alpha \\ \alpha^2 & -\alpha & 1 \end{pmatrix}$$

Which properties does $\mathbf{A}^{-1}$ have? Use the Cholesky decomposition to verify results.

Solution

$$\begin{aligned} \mathbf{A}^{-1} &= \frac{1}{(\alpha^2-1)^2} \begin{pmatrix} 1-\alpha^2 & \alpha^2(\alpha^2-1) & \phi \\ \alpha^2(\alpha^2-1) & (1-\alpha^2)(\alpha^2+1) & \alpha^2(\alpha^2-1) \\ \phi & \alpha^2(\alpha^2-1) & (1-\alpha^2) \end{pmatrix} \\ &= \frac{1}{(1-\alpha^2)} \begin{pmatrix} 1 & \alpha^2 & \phi \\ \alpha^2 & (\alpha^2+1) & \alpha^2 \\ \phi & \alpha^2 & 1 \end{pmatrix} \end{aligned}$$

Properties:Symmetric, Toeplitz and Hankel. Cholesky decomposition will yield the same result.

2.9 A real quadratic form is given by

$$g(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} - 2\mathbf{b}^T\mathbf{x} + c$$

where $\mathbf{A}$ is assumed to be a real, symmetric, positive definite matrix, $\mathbf{b}$ is a real $n \times 1$ vector, and c is a real constant. Solve for a minimum value of $g(\mathbf{x})$.

Solution

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} g(\mathbf{x}) &= 2\mathbf{A}\mathbf{x} - 2\mathbf{b} = \mathbf{0} \\ \Rightarrow & \quad \mathbf{x}_{opt} = \mathbf{A}^{-1}\mathbf{b} \\ \frac{\partial^2}{\partial \mathbf{x}^2} g(\mathbf{x}) &= 2\mathbf{A} \end{aligned}$$

Provided $\mathbf{A}$ is positive definite $\Rightarrow \mathbf{x}_{opt}$ is the minimum solution of $g(\mathbf{x})$.

2.10 Determine whether or not each of the following are valid autocorrelation matrices. If they are not, explain why not.

(a) $\mathbf{R_1} = \begin{bmatrix} 4 & 1 & 1 \\ -1 & 4 & 1 \\ -1 & -1 & 4 \end{bmatrix}$

(b) $\mathbf{R_2} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$

(c) $\mathbf{R_3} = \begin{bmatrix} 1 & 1+j \\ 1-j & 1 \end{bmatrix}$

(d) $\mathbf{R_4} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}$

(e) $\mathbf{R_5} = \begin{bmatrix} 2 & j & 1 \\ -j & 4j & -j \\ 1 & j & 2 \end{bmatrix}$

Solution

(a) Since $\mathbf{R_1}$ is not symmetric, it is *not* a valid autocorrelation matrix.

(b) Since $\mathbf{R_2}$ is symmetric and non-negative definite, this *is* a valid autocorrelation matrix.

(c) Although $\mathbf{R_3}$ is Hermitian, note that its determinant is negative,

$$\det \mathbf{R_3} = 1 - (1 + j)(1 - j) = -1$$

Therefore, $\mathbf{R_3}$ is not non-negative definite and, therefore, it is *not* a valid autocorrelation matrix.

(d) $\mathbf{R_4}$ *is* a valid autocorrelation matrix since it is symmetric and non-negative definite.

(e) The entries along the diagonal of an autocorrelation matrix must be real-valued (this follows from the Hermitian property, and the fact that the ith entry along the diagonal is equal to $E\{|x(i)|^2\}$ which is real). Since the middle element is imaginary, $\mathbf{R_5}$ is *not* a valid autocorrelation matrix.

Problem Set III: ARMA Modelling

Notice that the AR models used in these questions are in the form

$$x(n) + a_1 x(n-1) + \cdots + a_p x(n-p) = w(n)$$

We used the formulation

$$x(n) = a_1 x(n-1) + a_2 x(n-2) + \cdots + a_p x(n-p) + w(n)$$

in the lectures. Therefore, this will result in an opposite sign of the AR coefficients $\{a_1, \ldots, a_p\}$ and no other difference from the lectures.

3.1 For an AR(1) process with real parameter $a[1]$ and $\sigma_w^2 > 0$, calculate the mean and the variance of the process for $|a| < 1$ and $|a| \geq 1$.Where should the poles of AR filter lie for the process to be WSS?

Solution

$$x[n] \quad = \quad -a[1]x[n-1] + w[n] = \sum_{p=-\infty}^{n} h[n-p]w[p]$$

For $|a[1]| < 1$

$$\begin{aligned} h[n] &= (-a[1])^n \text{ for } n \geq 0 \\ E\{x[n]\} &= \sum_{p=-\infty}^{n} h[n-p]E\{w[p]\} = 0 \end{aligned}$$

Variance is given as $r_{xx}[0] = \frac{\sigma_w^2}{1-a^2[1]}$ For $|a[1] \geq 1$, $h[n] = (-a[1])^n$, constant for all n. Grows without bound for $|a[1]| > 1 \Rightarrow$ Infinite Variance output.

3.2 The autocorrelation function of the random process can be found from

$$\hat{r}_{xx}[n] = \frac{1}{2\pi j} \int_{|z|=1} \sigma_{xx}(z) z^{k-1} dz$$

Use this result to find the ACF of a real AR(1) process by evaluation of inverse z-transform of

$$\sigma_{xx}(z) = \frac{\sigma_w^2}{(1 + a[1]z^{-1})(1 + a[1]z)}$$

Assume $|a| < 1$.

Solution

$$r_{xx}[k] \quad = \quad \frac{1}{2\pi j} \int_{|z|=1} \frac{\sigma_w^2 z^k}{(z + a[1])(1 + a[1]z)} dz = \frac{1}{2\pi j} \times 2\pi j \times \text{Res at } z = -a[1] = \frac{\sigma_w^2 (-a[1])^k}{(1 - a^2[1])}$$

3.3 Use the same approach as in question 2.2 to verify that the ACF of the real AR(2) process with complex poles is given by

$$\hat{r}_{xx}[k] = \frac{\sigma_w^2 \frac{1+r^2}{1-r^2} \sqrt{1 + \left(\frac{1-r^2}{1+r^2}\right)^2 \cot^2(2\pi f_0)}}{1 - 2r^2 \cos(4\pi f_0) + r^4}$$

where

$$\psi = \tan^{-1} \left[\frac{1 - r^2}{1 + r^2} \cot(2\pi f_0)\right]$$

Solution

Method involves

$$r_{xx}[k] \;=\; \frac{1}{2\pi j}\int_{|z|=1}\frac{\sigma_w^2 z^{k+1}}{(z-\rho e^{j\theta})(z-\rho e^{-j\theta})(1-z\rho e^{j\theta})(1-z\rho e^{-j\theta})}dz$$

Poles within contour are at $z = \rho e^{+j\theta}$ and $z = \rho e^{-j\theta}$. Residues are complex conjugates of each other.
$r_{xx}[k] = $ Res @ $\rho e^{j\theta}$ + Res @ $\rho e^{-j\theta}$.
Residue at $\rho e^{j\theta}$ is $\frac{\sigma_N^2 \rho^k e^{jk\theta}}{2\sin(\theta)(1-\rho^2)j(e^{-j\theta}-\rho^2 e^{j\theta})}$ which requires:

$$\begin{aligned} &2\mathrm{Re}[\frac{e^{jk\theta}}{j(e^{-j\theta}-\rho^2 e^{j\theta})}] \\ &= 2\frac{(1+\rho^2)\sin(\theta)\cos k\theta + [1-\rho^2]\cos\theta\sin k\theta}{(1+\rho^2)^2\sin^2(\theta)+(1-\rho^2)^2\cos^2\theta}\end{aligned}$$

Residue at $\rho e^{j\theta}$ + Res at $\rho e^{-j\theta}$ is

$$\frac{\sigma_w^2\rho^k\frac{(1+\rho^2)}{(1-\rho^2)}(\cos k\theta + \frac{[1-\rho^2]}{[1+\rho^2]}\cot\theta\sin k\theta)}{(1-2\rho^2\cos(2\theta)+\rho^4)}$$

Hence, the result follows.

3.4 When signals are being observed by real world sensors, they are often corrupted by measurement noise $q[n]$ of variance σ_q^2. Consider the original signal $x[n]$ produced by an autoregressive process of order $p = 1$ (AR(1)), given by

$$x[n] = a_1 x[n-1] + w[n] \qquad\qquad w[n] \sim \mathcal{N}(0, \sigma_w^2)$$

which is measured as the noise corrupted process $y[n]$ given by

$$y[n] = x[n] + q[n] \qquad\qquad q[n] \sim \mathcal{N}(0, \sigma_q^2)$$

a) Show that the autoregressive parameter $\hat{a}_1$ is estimated from

$$\hat{a}_1 = \frac{r_{yy}(1)}{r_{yy}(0)}$$

where symbols $r_{yy}(1)$ and $r_{yy}(0)$ denote respectively the autocorrelations at lags $k = 1$ and $k = 0$.

b) Show that the relation between the true AR coefficient a_1 and the estimated AR coefficient $\hat{a}_1$ is

$$\hat{a}_1 = a_1\frac{\eta}{\eta+1}$$

where $\eta = r_{xx}(0)/\sigma_q^2$ is the signal to noise ratio (SNR).

c) Show that the noisy measurement $y[n]$ can be modelled as an ARMA(1,1) process, where the MA filter parameter b_1 is given as the solution of

$$\frac{1+b_1^2}{b_1} = \sigma_w^2 + \frac{\sigma_q^2(1+a_1^2)}{a_1\sigma_q^2}$$

Comment on the values of b_1 for a large ad small ratio σ_w^2/σ_q^2.

Hint: compare the power spectrum of P_{yy} and the power spectrum of an ARMA(1,1) process $P(z) = \frac{\left(1+b_1z^{-1}\right)\left(1+b_1z\right)}{\left(1+a_1z^{-1}\right)\left(1+a_1z\right)}$.

Solution

The true AR(1) coefficient a_1 is estimated from

$$a_1 = \frac{r_{xx}(1)}{r_{xx}(0)}$$

Since we can only observe the noisy measurement $y[n] = x[n] + q[n]$, the coefficient $\hat{a}_1$ is estimated from

$$\hat{a}_1 = \frac{r_{yy}(1)}{r_{yy}(0)}$$

that is, based on the noisy measurement $y[n]$.
b) Since the measurement noise $q[n]$ is white (e.g. $q[n] \perp q[n+1]$, $x \perp q$ and $E\{q^2(n)\} = \sigma_q^2$), we have

$$\begin{aligned} r_{yy}(0) &= E\{y[n]y[n]\} = E\{(x[n]+q[n])(x[n]+q[n])\} = r_{xx}(0) + \sigma_q^2 \\ r_{yy}(1) &= E\{y[n]y[n+1]\} = E\{(x[n]+q[n])(x[n+1]+q[n+1])\} = r_{xx}(1) \end{aligned}$$

Thus

$$\hat{a}_1 = \frac{r_{xx}(1)}{r_{xx}(0)+\sigma_q^2} = a_1 \frac{1}{1+\frac{1}{\eta}} = a_1 \frac{\eta}{1+\eta} \qquad \text{where} \qquad \eta = r_{xx}(0)/\sigma_q^2$$

Hence, as the SNR reduces, the coefficients of an AR model reduce in magnitude.
c) Since we have $y[n] = x[n] + q[n]$, then the corresponding power spectrum

$$\begin{aligned} P_{yy}(z) &= P_{xx}(z) + P_{qq}(z) = \frac{\sigma_w^2}{(1-a_1z^{-1})(1-a_1z}) + \sigma_q^2 \\ &= \frac{\sigma_w^2 + \sigma_q^2\big(1+a_1z^{-1}\big)\big(1+a_1z\big)}{\big(1+a_1z^{-1}\big)\big(1+a_1z\big)} \end{aligned}$$

Compare with the spectrum of ARMA(1,1) process

$$P(z) = \frac{\big(1+b_1z^{-1}\big)\big(1+b_1z\big)}{\big(1+a_1z^{-1}\big)\big(1+a_1z\big)}$$

and equate the numerator coefficients to yield

$$\begin{aligned} &(1) \qquad \sigma_w^2 + \sigma_q^2(1+a_1^2) = 1 + b_1^2 \\ &(2) \qquad \sigma_q^2 a_1(z+z^{-1}) = b_1(z+z^{-1}) \end{aligned}$$

to yield

$$\frac{\sigma_w^2 + \sigma_q^2(1+a_1^2)}{\sigma_q^2 a_1} = \frac{1+b_1^2}{b_1}$$

Therefore for a low SNR the ratio σ_w^2/σ_q^2 is small and $a_1 \to b_1$. For a high SNR $b_1 \to 0$ and we have the true AR signal.

3.5 Prove that the forward and backward AR processes

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + w[n]$$

and

$$x[n] = -\sum_{k=1}^{p} a[k]x[n+k] + w[n]$$

have the same Power Spectral Densities (PSDs).

Solution

Forward:

$$x[n] + \sum_{k=1}^{p} x[k]x[n-k] = w[n]$$

$$\text{Z tranform:} \quad X(z)\big[1 + a[1]z^{-1} + \cdots + a[p]z^{-p}\big] = W(z)$$

Backward:

$$x[n] + \sum_{k=1}^{p} x[k]x[n+k] = w[n]$$

$$\text{Z tranform:} \quad X(z)\big[1 + a[1]z^{1} + \cdots + a[p]z^{p}\big] = W(z)$$

$$P_{xx}(z) \quad = \quad \frac{\sigma_w^2}{A(z)A^*(\frac{1}{z^*})} = \frac{\sigma_w^2}{\Big[1 + a[1]z^{-1} + \cdots + a[p]z^{-p}\Big]\Big[1 + a[1]z^{1} + \cdots + a[p]z^{p}\Big]} = W(z)$$

which is independent of the forward or backward AR process.

3.6 Consider the complex MA(2) process

$$x[n] = w[n] - (3/2 + j)w[n-1] + 1/2(1+j)w[n-2]$$

Find the parameters of three other MA processes which have identical PSDs. Classify the phase response of each of the processes.

Solution

$$\begin{aligned} x[n] &= w[n] - (\frac{3}{2} + j)w[n-1] + \frac{1}{2}(1+j)w[n-2] \\ H_{MA} &= 1 - (\frac{3}{2} + j)z^{-1} + \frac{1}{2}(1+j)z^{-1} \\ &= (1 - (1+j)z^{-1})(1 - \frac{1}{2}z^{-1}) \end{aligned}$$

which corresponds to a mixed phase process.

$$H_{MA} \quad = \quad (1 - (\frac{1}{2} + \frac{j}{2})z^{-1})(1 - 2z^{-1}) = (1 - (\frac{5}{2} + \frac{j}{2})z^{-1}(1+j)z^{-2})$$

which is a 2nd mixed phase MA(2) process.

$$\begin{aligned} H_{MA} &= (1 - (\frac{1}{2} + \frac{j}{2})z^{-1})(1 - \frac{1}{2}z^{-1}) \\ &= (1 - (1 + \frac{j}{2})z^{-1}) + (\frac{1}{4}(1+j)z^{-2}) \end{aligned}$$

which is a minimum phase MA(2) process.

$$\begin{aligned} H_{MA} &= (1 - (1+j)z^{-1})(1 - 2z^{-1}) \\ &= (1 - (3+j)z^{-1} + 2(1+j)z^{-1}) \end{aligned}$$

which is a maximum phase solution.

3.7 In some applications, the data collection process may be flawed so that there are either missing data values or outliers should be discarded. Suppose that we are given N samples of a WSS process $x(n)$ with one value, $x(n_0)$, missing. Let **x** be the vector containing the given sample values,

$$\mathbf{x} = [x(0), x(1), \dots, x(n_0 - 1), x(n_0 + 1), \dots, x(N)]^T$$

(a) Let $\mathbf{R}_x$ be the autocorrelation matrix for the vector **x**,

$$\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\}$$

Which of the following statements are true:

1. $\mathbf{R}_x$ is Toeplitz.
2. $\mathbf{R}_x$ is Hermitian.
3. $\mathbf{R}_x$ is Positive semidefinite.

(b) Given the autocorrelation matrix for **x**, is it possible to find the autocorrelation matrix for the vector

$$\mathbf{x} = [x(0), x(1), \dots, x(N)]^T$$

that does not have $x(n_0)$ missing? If so, how would you find it? If not, explain why not.

Solution

(a) The matrix is *not Toeplitz*. This may be shown easily by example. If **x** = $[x(0), x(2), x(3)]$, then

$$\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\} = \begin{bmatrix} |x(0)|^2 & x(0)x^*(2) & x(0)x^*(3) \\ x(2)x^*(0) & |x(2)|^2 & x(2)x^*(3) \\ x(3)x^*(0) & x(3)x^*(2) & |x(3)|^2 \end{bmatrix} = \begin{bmatrix} r_x(0) & r_x(2) & r_x(3) \\ r_x(2) & r_x(0) & r_x(1) \\ r_x(3) & r_x(1) & r_x(0) \end{bmatrix}$$

which is clearly not Toeplitz. However, by definition, $\mathbf{R}_x$ is Hermitian,

$$\mathbf{R}_x^H = E\{\mathbf{x}\mathbf{x}^H\}^H = E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{R}_x$$

Finally, $\mathbf{R}_x$ is non-negative definite, which may be shown as follows. Let **v** be any non-zero vector. Then,

$$\mathbf{v}^H\mathbf{R}_x\mathbf{v} = \mathbf{v}^H E\{\mathbf{x}\mathbf{x}^H\}\mathbf{v} = E\{\mathbf{v}^H\mathbf{x}\mathbf{x}^H\mathbf{v}\}$$

Therefore,

$$\mathbf{v}^H\mathbf{R}_x\mathbf{v} = E\{|\mathbf{v}^H\mathbf{x}|^2\} \geqslant 0$$

and $\mathbf{R}_x \geqslant 0$.

(b) There are several ways to find the autocorrelation matrix for the vector

$$\mathbf{x} = [x(0), x(1), \dots, x(N)]^T$$

that does not have $x(n_0)$ missing. One way is as follows. Note that the first column of $\mathbf{R}_x$ that is formed from the vector that has $x(n_0)$ missing is as follows,

$$[r_x(0), r_x(1), \dots, r_x(n_0 - 1), r_x(n_0 + 1), \dots, r_x(N)]$$

Therefore, all that we need is the missing correlation $r_x(n_0)$. Note, however, that this term is found in the second column of row (n_0+2) (see example in part (a) above). Thus, given $r_x(n_0)$ the Toeplitz matrix may then be formed.

3.8 Voiced speech may be modeled as the output of an all-pole filter driven by an impulse train

$$p_{n_0}(n) = \sum_{k=1}^{K} \delta(n - kn_0)$$

where the time between pulses, n_0, is known as the *pitch period*. Suppose that we have a segment of voiced speech, and that we know the pitch period, n_0. We extract a subsequence, $x(n)$, of length $N = 2n_0$ and model this signal as shown in the following figure

$$\xrightarrow{p_{n_0}(n)} \boxed{\frac{b(0)}{1+\Sigma_{k=1}^{p} a_p(k) z^{-k}}} \xrightarrow{\widehat{x}(n)}$$

where the input, $p_{n_0}(n)$, consists of two pulses,

$$p_{n_0}(n) = \delta(n) + \delta(n - n_0)$$

Find the normal equations that define coefficients $a_p(k)$ that minimize the error

$$\varepsilon_p = \sum_{n=0}^{N-1} e^2(n)$$

where

$$e(n) = a_p(n) * x(n) - b(n) * p_{n_0}(n)$$

and $b(n) = b(0)\delta(n)$.

Solution

If we define $a_p(0) = 1$, then the error $e(n)$ is

$$e(n) = a_p(n) * x(n) - b(n) * p_{n_0}(n) = \sum_{l=0}^{p} a_p(l) x(n-l) - b(0)[\delta(n) + \delta(n - n_0)]$$

and the mean-square error that we want to minimize is

$$\varepsilon_p = \sum_{n=0}^{2n_0-1} e^2(n) = \sum_{n=0}^{2n_0-1} \left[\sum_{l=0}^{p} a_p(l) x(n-l) - b(0)\delta(n) - b(0)\delta(n-n_0) \right]^2 \tag{15}$$

Setting the derivative with respect to $a_p(k)$ equal to zero, we have

$$\frac{\partial \varepsilon_p}{\partial a_p(k)} = \sum_{n=0}^{2n_0-1} 2 \left[\sum_{l=0}^{p} a_p(l) x(n-l) - b(0)\delta(n) - b(0)\delta(n-n_0) \right] x(n-k) = 0$$

If we define

$$r_x(k,l) = \sum_{n=0}^{2n_0-1} x(n-l) x(n-k)$$

then the normal equations become (recall that $a_p(0) = 1$)

$$\sum_{l=1}^{p} a_p(l) r_x(k,l) - b(0) x(-k) - b(0) x(n_0 - k) = -r_x(k,0); \;\; k = 1, 2, \ldots, p$$

Assuming that $x(n) = 0$ for n<0, with $\mathbf{x} = [x(n_0-1), x(n_0-2), \ldots, x(n_0-p)]^T$, the normal equations may be written in matrix form as follows

$$\mathbf{R}_x \mathbf{a} - b(0)\mathbf{x} = -\mathbf{r}_x$$

Finally, differentiating with respect to $b(0)$ we have

$$\frac{\partial \varepsilon}{\partial b(0)} = -\sum_{n=0}^{\infty} 2 \left[\sum_{l=0}^{p} a_p(l)x(n-l) - b(0)\delta(n) - b(0)\delta(n-n_0) \right] [\delta(n) + \delta(n-n_0)]$$

Thus,

$$x(0) - b(0) + \sum_{l=1}^{p} a_p(l)x(n_0 - l) - b(0) = -x(n_0)$$

or, in vector form, we have

$$\mathbf{x}^T \mathbf{a} - 2b(0) = -x(0) - x(n_0)$$

Putting all of these together in matrix form yields

$$\begin{bmatrix} \mathbf{R}_x & \mathbf{x} \\ \mathbf{x}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ -2b(0) \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_x \\ x(0) + x(n_0) \end{bmatrix}$$

Problem Set IV: Spectrum Estimation

4.1 Evaluate the mean and the autocorrelation of the sequence $x[n]$ which is generated by the MA(2) process described by

$$x[n] = w[n] - 2w[n-1] + w[n-2]$$

where $w[n]$ is the white noise process of variance σ_w^2.

Solution

$$\begin{aligned} E\{x[n]\} &= E\{w[n]\} - 2E\{w[n-1]\} + E\{w[n-2]\} \\ &= 0, \qquad \text{with the assumption } E\{w[n]\} = 0 \end{aligned}$$

$$r_{xx}(0) = E\{(w[n] - 2w[n-1] + w[n-2])^2\} = \sigma_w^2(1+4+1) = 6\sigma_w^2$$

$$r_{xx}(1) = E\{(w[n] - 2w[n-1] + w[n-2])(w[n+1] - 2w[n] + w[n-1])\} = -4\sigma_w^2$$

$$r_{xx}(2) = E\{(w[n] - 2w[n-1] + w[n-2])(w[n+2] - 2w[n+1] + w[n])\} = \sigma_w^2$$

$$\begin{aligned} r_{xx}(k) &= r_{xx}(-k) \quad \text{for } k = 1, 2 \\ r_{xx}(k) &= 0 \quad \text{for } k \geq 3 \end{aligned}$$

4.2 Consider the AR(2) process generated by the equation

$$x[n] = 14/24x[n-1] + 9/24x[n-2] - 1/24x[n-3] + w[n]$$

where $w[n]$ is the stationary white noise process of variance σ_w^2

(a) Calculate the coefficients of the optimum $p = 3$ linear predictor.

(b) Determine the correlation sequence $r_{xx}[m]$, $0 \leq m \leq 5$.

(c) **Optional:** Determine the reflection coefficients which correspond to $p = 3$ linear predictor.

Solution

$$A(z) = (1 - 14/24z^{-1} - 9/24z^{-2} + 1/24z^{-3})$$

Use the step down algorithm:

$$\begin{aligned} \rho_3 &= \sigma_w^2, \rho_2 = 1.002\sigma_w^2, \rho_1 = 1.143\sigma_w^2, \rho_0 = 4.94\sigma_w^2, \\ a_3[3] &= 1/24, a_2[2] = -0.351, a_1[1] = -0.877 \\ a_3[2] &= -9/24, a_2[1] = -0.569, a_3[1] = -14/24 \end{aligned}$$

Now $r_{xx}(0) = \rho_0 = 4.94\sigma_w^2$. Next

$$\begin{aligned} a_1[1] &= \frac{-r_{xx}(1)}{r_{xx}(0)} = -0.877 = \frac{-r_{xx}(1)}{4.94\sigma_w^2} \\ \Rightarrow r_{xx}(1) &= 4.332\sigma_w^2 \end{aligned}$$

$$\begin{aligned} a_2[2] &= -(\frac{r_{xx}(2)+a_1[1]r_{xx}(1)}{\rho_1}) = -(\frac{r_{xx}(2)+(-0.877)(4.332\sigma_W^2)}{1.143\sigma_w^2}) \\ \Rightarrow r_{xx}(2) &= 4.2\sigma_w^2 \end{aligned}$$

$$\begin{aligned} a_3[3] &= -(\frac{r_{xx}(3)+a_2[1]r_{xx}(2)+a_2[2]r_{xx}(1)}{\rho_2}) \\ \Rightarrow r_{xx}(3) &= 3.869\sigma_w^2 \end{aligned}$$

>From the extrapolation of the ACF:

$$\begin{aligned} -r_{xx}[4] &= a_3[1]r_{xx}[3]+a_3[2]r_{xx}[2]+a_3[3]r_{xx}[1] \\ &= \frac{-14}{24}3.869\sigma_w^2 - \frac{9}{24}4.2\sigma_w^2 + \frac{1}{24}4.332\sigma_w^2 \\ r_{xx}(4) &= 3.651\sigma_w^2 \end{aligned}$$

Lastly,

$$\begin{aligned} -r_{xx}[5] &= a_3[1]r_{xx}[4]+a_3[2]r_{xx}[3]+a_3[3]r_{xx}[2] \\ &= \frac{-14}{24}3.651\sigma_w^2 - \frac{9}{24}43.869\sigma_w^2 + \frac{1}{24}4.2\sigma_w^2 \\ r_{xx}(5) &= 3.406\sigma_w^2 \end{aligned}$$

>From Part 2:

$$\Gamma_1 = -0.877$$
$$\Gamma_2 = -0.351$$
$$\Gamma_3 = -1/24$$

4.3 (a) Evaluate the power coefficients of the random processes which are generated by

$$\begin{aligned} &(1) \qquad x[n] = -0.81x[n-2]+w[n]-w[n-1] \\ &(2) \qquad x[n] = w[n]-w[n-2] \\ &(3) \qquad x[n] = -0.81x[n-2]+w[n] \end{aligned}$$

where $w[n]$ is the stationary white noise process of variance σ_w^2

(b) Sketch the spectra of the process given in part (a).

(c) Determine the autocorrelation function $r_{xx}[m]$ for the processes (2) and (3).

Solution

(a) (1) ARMA(2,1):

$$H(z) = \frac{1-z^{-1}}{1+0.81z^{-1}}$$

$$P_{ARMA}(f) = \sigma_w^2 \left|\frac{1-e^{-j2\pi f}}{1+0.81e^{-j2\pi f}}\right|^2$$

(2) - MA(2):

$$H(z) = 1-z^{-2}$$

$$P_{MA}(f) = \sigma_w^2 \left|1-e^{-j4\pi f}\right|^2$$

(3) - AR(2):

$$H(z) = \frac{1}{1+0.81z^{-1}}$$

$$P_{AR}(f) = \sigma_w^2 \left| \frac{1}{1+0.81e^{-j2\pi f}} \right|^2$$

(b):

$$\begin{aligned} 2\sigma_w^2 &= r_{xx}(0) \\ 0 &= r_{xx}(1) = r_{xx}(-1) \\ -\sigma_w^2 &= r_{xx}(2) = r_{xx}(-2) \end{aligned}$$

(c)

$$r_{xx}(k) = \frac{\sigma_w^2(-0.81)^{|k|}}{(1-(0.81)^2)}$$

4.4 Bias Analysis of the Periodogram for different Windows. The expected value of the spectrum estimate via the DTFT of the ACF of a signal using a window w can be written as

$$E\left\{\hat{\phi}(\omega)\right\} = \sum_{k=-\infty}^{\infty} [w(k)r(k)]e^{-j\omega k}. \tag{16}$$

As the DTFT of the product of two sequence is the convolution of their respective DTFTs, Equation (**??**) leads to

$$E\left\{\hat{\phi}(\omega)\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(\psi)W(\omega-\psi)d\psi. \tag{17}$$

Show the explicit mathematical expression of $W(\omega)$ for both triangular and rectangular windows. Provide plots for both functions.

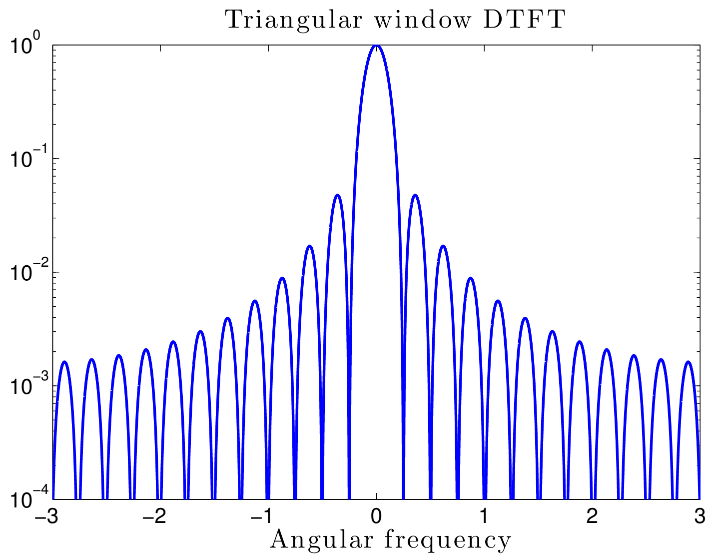
Solution

- Triangular window: The expression of the triangular window DTFT is given by

$$\begin{aligned} W_T(\omega) &= \sum_{k=-(N-1)}^{N-1} \frac{N-|k|}{N} e^{-j\omega k} \qquad (18) \\ &= \frac{1}{N}\sum_{n=1}^{N}\sum_{s=1}^{N} e^{-j\omega(n-s)} = \frac{1}{N}\left|\sum_{n=1}^{N} e^{j\omega n}\right|^2 \\ &= \frac{1}{N}\left|\frac{e^{j\omega N}-1}{e^{j\omega}-1}\right|^2 = \frac{1}{N}\left|\frac{e^{j\omega N/2}-e^{-j\omega N/2}}{e^{j\omega/2}-e^{-j\omega/2}}\right|^2, \end{aligned}$$

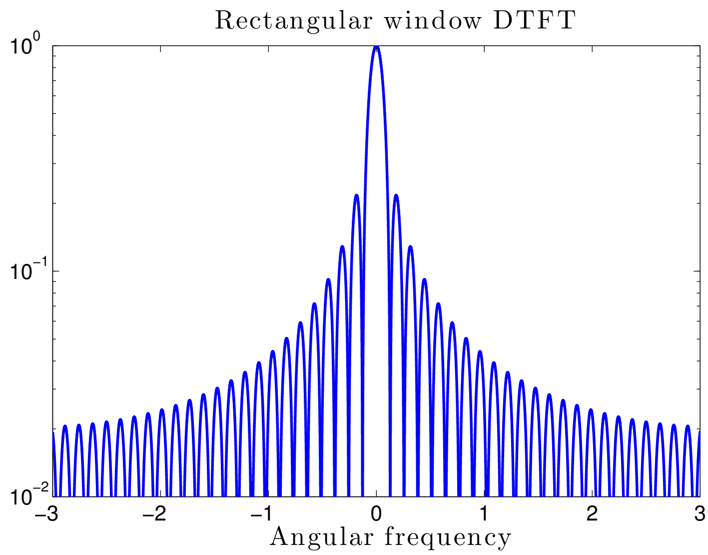
or in the final form

$$W_T(\omega) = \frac{1}{N}\left[\frac{\sin(\omega N/2)}{\sin(\omega/2)}\right]^2. \tag{19}$$



- Rectangular window: The expression of the rectangular window DTFT is given by

$$
\begin{aligned}
W_R(\omega) &= \sum_{k=-(N-1)}^{N-1} e^{-j\omega k} = \Re\left[\frac{e^{jN\omega}-1}{e^{j\omega}-1}\right] - 1 \\
&= \frac{2\cos\left[\frac{(N-1)\omega}{2}\right]\sin\left[\frac{N\omega}{2}\right]}{\sin\left[\frac{\omega}{2}\right]} - 1 = \frac{\sin\left[\left(N-\frac{1}{2}\right)\omega\right]}{\sin\left[\frac{\omega}{2}\right]}
\end{aligned}
\tag{20}
$$



Note that from Eqs. (**??**) and (**??**) we can estimate the main lobe width by determining the values of ω for which the argument of the numerator lies within the interval $[-\pi, \pi]$, this way:

$$
\begin{array}{rcll}
-\pi \leq & \frac{\omega N}{2} & \leq \pi & \text{(Triangular window)} \\
-\pi \leq & \left(N-\frac{1}{2}\right)\omega & \leq \pi & \text{(Rectangular window)}
\end{array}
$$

This way, for $N >> 1$ we the main lobe widths of both the triangular and rectangular windows are respectively given by

$$
w_T \simeq \frac{2\pi}{N}, \quad w_R = \frac{4\pi}{N}. \tag{21}
$$

4.5 The harmonic decomposition problem considered by Pisarenko may be expressed as the solution to the equation

$$
\mathbf{a}^H \mathbf{R}_{yy} \mathbf{a} = \sigma_w^2 \mathbf{a}^H \mathbf{a}
$$

The solution for $\mathbf{a}$ may be obtained by minimization of the quadratic form $\mathbf{a}^H\mathbf{R}_{yy}\mathbf{a}$ subject to the constraint $\mathbf{a}^H\mathbf{a} = 1$. the constraint can be incorporated to the index with the Lagrange multiplier. Thus the performance index becomes

$$E = \mathbf{a}^H\mathbf{R}_{yy}\mathbf{a} + \lambda(1 - \mathbf{a}^H\mathbf{a})$$

By minimizing E with respect to $\mathbf{a}$ show that this formulation is equivalent to the Pisarenko eigenvalue problem with the Lagrange multiplier λ playing the role of the eigenvalue. Thus, show that the minimization of E is the minimum eigenvalue of $\mathbf{R}_{yy}$.

Solution

$$E = \mathbf{a}^H\mathbf{R}_{yy}\mathbf{a} + \lambda(1 - \mathbf{a}^H\mathbf{a})$$

$$E = \mathbf{a}^H\mathbf{R}_{yy}\mathbf{a} + \lambda(1 - \mathbf{a}^H\mathbf{a})$$

$$\begin{aligned}\nabla_{\mathbf{a}}E = \frac{\partial}{\partial\mathbf{a}}E = 2\mathbf{R}_{yy}\mathbf{a} - 2\lambda\mathbf{a} = \mathbf{0}\\ \Rightarrow (\mathbf{R}_{yy} - \lambda\mathbf{I})\mathbf{a} = \mathbf{0}\end{aligned}$$

Hence, eigen-equation with $\lambda \Leftrightarrow$ eigenvalues

$$\nabla_{\mathbf{a}}\big(\nabla_{\mathbf{a}}E\big)^T = \frac{\partial^2}{\partial\mathbf{a}^2}E = \mathbf{R}_{yy} - \lambda\mathbf{I}$$

The above matrix is positive definite when $\lambda = \lambda_{MIN}$.

4.6 The autocorrelation of a sequence which consists of a sinusoid with random phase in noise is

$$r_{xx}[m] = P\cos 2\pi f_1 m + \sigma_w^2$$

where f_1 is the frequency of the sinusoid, P its power, and σ_w^2 the variance of noise. Consider fitting an AR(2) model to the data.

(a) Determine the optimum coefficients of the AR(2) model as a function of σ_w^2 and f_1.

(b) **Optional:** Determine the reflection coefficients corresponding to the AR(2) model parameters.

(c) Determine the limiting values of the AR(2) model parameters and (K_1, K_2) as $\sigma_w^2 \to 0$.

Solution

$$\begin{pmatrix} P+\sigma_w^2 & P\cos(2\pi f_1) \\ P\cos(2\pi f_1) & P+\sigma_w^2 \end{pmatrix} \cdot \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \end{pmatrix} = -\begin{pmatrix} P\cos(2\pi f_1) \\ P\cos(4\pi f_1) \end{pmatrix}$$

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \end{pmatrix} = \frac{1}{P^2\cos^2(2\pi f_1) - (P+\sigma_w^2)^2} \begin{pmatrix} P+\sigma_w^2 & -P\cos(2\pi f_1) \\ -P\cos(2\pi f_1) & P+\sigma_w^2 \end{pmatrix} \cdot \begin{pmatrix} P\cos(2\pi f_1) \\ P\cos(4\pi f_1) \end{pmatrix}$$

(a).

$$\hat{a}_1 = \frac{P\cos(2\pi f_1)(P+\sigma_w^2 - P\cos(4\pi f_1))}{P^2\cos^2(2\pi f_1) - (P+\sigma_w^2)^2}$$

$$\hat{a}_2 = \frac{P([P+\sigma_w^2]\cos(4\pi f_1) - P\cos^2(2\pi f_1))}{P^2\cos^2(2\pi f_1) - (P+\sigma_w^2)^2}$$

(b).

$$\kappa_1 = \Gamma_1 = \frac{-r_{xx}[1]}{r_{xx}[0]} = \frac{-P\cos(2\pi f_1)}{P + \sigma_w^2}$$

$$\kappa_2 = \Gamma_2 = \hat{a}_2 = \frac{P([P + \sigma_w^2]\cos(4\pi f_1) - P\cos^2(2\pi f_1))}{P^2\cos^2(2\pi f_1) - (P + \sigma_w^2)^2}$$

(c).

$$\begin{aligned} \hat{a}_1 &\to -\cos(2\pi f_1) \\ \hat{a}_2 &\to 1 \\ \Gamma_1 &\to -\cos(2\pi f_1) \\ \Gamma_2 &\to 1 \end{aligned}$$

4.7 Given $N = 10,000$ samples of a process $x(n)$, you are asked to compute the periodogram. However, with only a finite amount of memory resources, you are unable to compute a DFT any longer than 1024. Using these $10,000$ samples, describe how you would be able to compute a periodogram that has a resolution of

$$\Delta\omega = 0.89\frac{2\pi}{10000}$$

Hint: Consider how the decimation-in-time FFT algorithm works.

Solution

To get the maximum resolution from $N = 10,000$ data values, we want to compute the periodogram of $x(n)$ (segmenting $x(n)$ into subsequences reduces the resolution). The question, therefore, is how to compute the periodogram of $x(n)$ using 1024-point DFT's. Recalling how the FFT works, note that

$$X(e^{j\omega}) = \sum_{n=0}^{9999} x(n)e^{-jn\omega} = \sum_{n=0}^{9999}\sum_{l=0}^{9} x(10n+l)e^{-j(10n+l)\omega} = \sum_{l=0}^{9} e^{-jl\omega}\sum_{n=0}^{9999} x(10n+l)e^{-jn\omega}$$

Therefore, the procedure is to pad $x(n)$ to form a sequence of length $N = 10240$, and then decimate $x(n)$ into 10 sequences $x_l(n)$ of length $M = 1024$,

$$x_l(n) = x(10n+l) \;\; ; \;\; n = 0, 1, \ldots, 1023$$

Next, the 1024-point DFT's of these sequences, $X_l(k)$, are computed, and combined using the "twiddle factors" $\exp(-jl\frac{2\pi k}{10240})$ as follows

$$X(k) = \sum_{l=0}^{9} e^{-jl\frac{2\pi k}{10240}} X_l(k) \;\; ; \;\; k = 0, 1, \ldots, 10239$$

Finally, squaring the magnitude of $X(k)$ and dividing by $N = 10240$, we have the periodogram with a resolution of $\Delta\omega = 0.89\frac{2\pi}{10000}$.

4.8 Given $N = 1000$ samples of the process:

$$x(n) = 2\cos(2\pi 100n) + \cos(2\pi 400n) + 2\cos(2\pi 410n) + \eta(n), \quad \eta \sim \mathcal{N}(0, 0.2),$$

you are asked to compute the least squares (LS) periodogram using 500 equally-spaced points for the frequency range between 2π and 1000π (that is, between 1Hz and 500Hz).

Solution

The LS periodogram can be computed based on the solution of:

$$\min_{\beta,\phi} \sum_{n=1}^{N} |x(n) - |\beta| \cos(\omega n + \phi)|^2, \tag{22}$$

or its reparametrised version ($a(\omega) = \beta(\omega)\cos\phi, b(\omega) = \beta(\omega)\sin\phi$):

$$\min_{a,b} \sum_{n=1}^{N} |x(n) - a\cos(\omega n) - b\sin(\omega n)|^2. \tag{23}$$

The solution to (**??**) is well known and given by:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \mathbf{R}^{-1}\mathbf{r}, \tag{24}$$

where:

$$\mathbf{R} = \sum_{n=1}^{N} \begin{bmatrix} \cos(wn) \\ \sin(wn) \end{bmatrix} \begin{bmatrix} \cos(wn) & \sin(wn) \end{bmatrix}, \quad \mathbf{r} = \sum_{n=1}^{N} \begin{bmatrix} \cos(wn) \\ \sin(wn) \end{bmatrix} x(n). \tag{25}$$

Then, the signal which solves (**??**) for frequency ω in an LS sense is given by $\hat{x} = a\cos(\omega n) + b\sin(\omega n)$ and its power is computed according to:

$$\begin{aligned} \hat{P}_\omega &= \frac{1}{N}\sum_{n=1}^{N}\Big(\hat{a}\sin(\omega n) + \hat{b}\cos(\omega n)\Big)^2 \qquad (26) \\ &= \frac{1}{N}\sum_{n=1}^{N}\Big(\hat{a}^2\sin^2(\omega n) + 2\hat{a}\hat{b}\sin(\omega n)\cos(\omega n) + \hat{b}^2\cos^2(\omega n)\Big) \\ &= \frac{1}{N}\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}^T \left(\sum_{n=1}^{N}\begin{bmatrix} \cos(wn) \\ \sin(wn) \end{bmatrix}\begin{bmatrix} \cos(wn) & \sin(wn) \end{bmatrix}\right)\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}. \end{aligned}$$

By replacing eqs. (**??**) and (**??**) in (**??**) we have:

$$\hat{P}_\omega = \frac{1}{N}\mathbf{r}^T\mathbf{R}^{-1}\mathbf{r}, \tag{27}$$

and consequently, the computation of the LS periodogram is achieved by evaluating (**??**) for frequencies of interest, i.e.

$$\hat{P}_{LS}(\omega) = \frac{1}{N}\mathbf{r}^T(\omega)\mathbf{R}^{-1}(\omega)\mathbf{r}(\omega), \qquad \omega \in [2\pi, 4\pi, ..., 2*500\pi]. \tag{28}$$

4.9 From measurements of a process $x(n)$, we estimate the following values for the autocorrelation sequence:

$$r_x(k) = \alpha^{|k|} \;\; ; \;\; |k| \leq M$$

where $|\alpha| < 1$. Estimate the power spectrum using

(a) The Blackman-Tukey method with a rectangular window.

(b) The minimum variance method.

(c) The maximum entropy method.

Solution

(a) Using the Blackman-Tukey method with a rectangular window we have

$$\begin{aligned} P_{BT}(e^{j\omega}) &= \sum_{k=-M}^{M} \alpha^{|k|} e^{-jk\omega} \\ &= \sum_{k=0}^{M} \alpha^k e^{-jk\omega} + \sum_{k=0}^{M} \alpha^k e^{jk\omega} - 1 \\ &= 2\alpha^{M+1} \frac{1-\alpha^2 + \cos(M\omega) - \cos(M+1)\omega}{1+\alpha^2 - 2\alpha\cos\omega} \end{aligned}$$

(Note: the last step above requires a bit of algebra).

(b) For the minimum variance method,

$$\widehat{P}_{MV}(e^{j\omega}) = \frac{M+1}{\mathbf{e}^H \mathbf{R}_M^{-1} \mathbf{e}} = \frac{M+1}{q(0) + 2\sum_{k=1}^{M} q(k)\cos k\omega}$$

where

$$\mathbf{R}_M = \text{Toep}\{r_x(0), r_x(1), \ldots, r_x(M)\}$$

and the coefficients $q(k)$ are the sums along the diagonals of $\mathbf{R}_M^{-1}$. As we saw in Example 5.2.11 (p. 258), the inverse of this autocorrelation matrix is

$$\mathbf{R}_M^{-1} = \frac{1}{1-\alpha^2} \begin{bmatrix} 1 & -\alpha & 0 & \ldots & 0 & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \ldots & 0 & 0 \\ 0 & -\alpha & 1+\alpha^2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1+\alpha^2 & -\alpha \\ 0 & 0 & 0 & \ldots & -\alpha & 1 \end{bmatrix}$$

Therefore,

$$\widehat{P}_{MV}(e^{j\omega}) = \frac{(M+1)(1-\alpha^2)}{(M+1) + (M-1)\alpha^2 - 2M\alpha\cos\omega}$$

(c) The maximum entropy spectrum is given by

$$\widehat{P}_{mem}(e^{j\omega}) = \frac{\epsilon_p}{\left|1 + \sum_{k=1}^{p} a_p(k) e^{jk\omega}\right|^2}$$

Solving the autocorrelation normal equations for the coefficients $a_p(k)$, we find

$$\mathbf{a}_p = [1, -\alpha, 0, \ldots, 0]^T$$

and

$$\epsilon_p = 1 - \alpha^2$$

Therefore, the MEM spectrum is

$$\widehat{P}_{mem}(e^{j\omega}) = \frac{1-\alpha^2}{|1-\alpha e^{-j\omega}|^2} = \frac{1-\alpha^2}{1+\alpha^2 - 2\alpha\cos\omega}$$

4.10 Nonnegativeness of the Blackman-Tukey spectral estimate.

a) Show that if the lag window $\{w(k)\}$ is positive semidefinite, that is, its spectrum $W(\omega) \geq 0$, then the windowed covariance sequence $\{w(k)\hat{r}(k)\}$, where

$$\hat{r}(k) = \frac{1}{N}\sum_{n=0}^{N-1-k} y(n)y(n+k), \qquad 0 \le k \le N-1$$

is positive definite too, thus implying $\hat{P}_{wx}(\omega) \ge 0$. Notice that the above standard autocorrelation estimator $\hat{r}$ is biased (asymptotically unbiased for a fast decaying ACF and large N).

b) (Optional) Prove that the Blackman-Tukey spectral estimate is positive semidefinite.

Solution

We will prove this in two ways. The part a) is valid for real and symmetric window (around point $k = 0$).

a) For a real and symmetric sequence $\{w(k)\}$, its Discrete Fourier Transform $W(\omega)$ is also an even and real function. Furthermore, if $\{w(k)\}$ is a positive semidefinite sequence, then $W(\omega) \ge 0$, for all ω. This immediately implies that $\hat{P}_{wx}(\omega) \ge 0$, as the power spectrum $\hat{P}_x$ is positive by definition.

It should be noted that some lag windows, such as the rectangular window, do not satisfy the positive semidefinite assumption, and they can lead to estimate spectra that have negative values.
The Bartlett window, on the other hand is positive semidefinite, as $W_B(\omega) = \frac{1}{N}\left[\frac{\sin(\omega N/2)}{\sin(\omega/2)}\right]^2 > 0$.
b) Recall that the Blackman-Tukey spectral estimate uses M out of the available N points of the autocorrelation estimate ($M < N$). Since the power spectrum is always positive semidefinite by definition (square of the DFT of a signal), then the Blackman-Tukey power spectrum estimate $P_{BT}(\omega)$ is positive semidefinite only if the sequence

$$\{\ldots, 0, w(-(M-1))\hat{r}(-(M-1)), \ldots, w(M-1)\hat{r}(M-1), 0, \ldots\}$$

is posidife semidefinite, or equivalently, the following Toeplitz matrix is positive semidefinite

$$\begin{bmatrix} w(0)\hat{r}(0) & \cdots & w(M-1)\hat{r}(M-1) & 0 \\ \vdots & \ddots & \vdots & \ddots \\ w(-(M-1))\hat{r}(-(M-1)) & \cdots & w(0)\hat{r}(0) & \cdots \\ & \ddots & \ddots & \\ 0 & & & \end{bmatrix} =$$

$$\begin{bmatrix} w(0) & \cdots & w(M-1) & 0 \\ \vdots & \ddots & \vdots & \ddots \\ w(-(M-1)) & \cdots & w(0) & \cdots \\ & \ddots & \ddots & \\ 0 & & & \end{bmatrix} \circ \begin{bmatrix} \hat{r}(0) & \cdots & \hat{r}(N-1) & 0 \\ \vdots & \ddots & \vdots & \ddots \\ \hat{r}(-(N-1)) & \cdots & \hat{r}(0) & \cdots \\ & \ddots & \ddots & \\ 0 & & & \end{bmatrix}$$

where the symbol '$\circ$' denotes the element-wise (Hadamard) product of matrices. By a result in matrix theory, the Hadamard product of two positive semidefinite matrices is also a positive definite matrix, which concludes the proof.

4.11 In the MUSIC algorithm, finding the peaks of the frequency estimation function

$$\widehat{P}_{MU}(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^{M} |\mathbf{e}^H \mathbf{v}_i|^2}$$

is equivalent to finding the minima of the denominator. Show that finding the *minima* of the denominator is equivalent to finding the *maxima* of

$$\sum_{i=1}^{p} |\mathbf{e}^H \mathbf{v}_i|^2$$

Hint: Use the fact that

$$\mathbf{I} = \sum_{i=1}^{M} \mathbf{v}_i \mathbf{v}_i^H$$

Solution

Finding the peaks of the MUSIC frequency estimation function is equivalent to finding the minima of

$$\sum_{i=p+1}^{M} |\mathbf{e}^H \mathbf{v}_i|^2$$

Since the eigenvectors $\mathbf{v}_i$ are orthogonal, if we assume that they have been normalized, then the identity matrix may be expanded in terms of these eigenvectors as follows,

$$\mathbf{I} = \sum_{i=1}^{M} \mathbf{v}_i \mathbf{v}_i^H$$

Multiplying on the left by $\mathbf{e}^H$ and on the right by $\mathbf{e}$, we have

$$\mathbf{e}^H \mathbf{e} = \sum_{i=1}^{M} (\mathbf{e}^H \mathbf{v}_i)(\mathbf{v}_i^H \mathbf{e}) = \sum_{i=1}^{M} |\mathbf{e}^H \mathbf{v}_i|^2$$

Since $\mathbf{e}^H \mathbf{e} = M$, then

$$M = \sum_{i=1}^{p} |\mathbf{e}^H \mathbf{v}_i|^2 + \sum_{i=p+1}^{M} |\mathbf{e}^H \mathbf{v}_i|^2 \qquad \text{or} \qquad \sum_{i=p+1}^{M} |\mathbf{e}^H \mathbf{v}_i|^2 = M - \sum_{i=1}^{p} |\mathbf{e}^H \mathbf{v}_i|^2$$

Thus, minimizing the left-hand side is equivalent to maximizing the sum on the right as was to be shown.

Problem Set V: Adaptive Filters

5.1 Derive the least mean square algorithm based on minimizing the instantaneous error (w.r.t $\mathbf{w}[n]$)

$$J[n] = e^2[n]$$

where

$$e[n] = d[n]\mathbf{w}^T[n] - \mathbf{x}[n]$$

and show, explicitly the update equation for the coefficients of the FIR filter.

Optional: Derive the equivalent LMS lattice algorithm for minimising the joint forward and backward error (w.r.t the parameter $\tilde{\gamma}[n]$)

$$J[n] = e_f^2[n] + e_b^2[n].$$

where

$$\begin{aligned} e_f[n] &= x[n] + \Gamma[n]x[n-1] \\ e_b[n] &= x[n-1] + \Gamma[n]x[n] \end{aligned} \tag{29}$$

Solution ____

LMS:

$$\begin{aligned} \mathbf{w}[n] &= \mathbf{w}[n-1] - \mu\frac{\partial}{\partial\mathbf{w}}J[n]_{|\mathbf{w}=\mathbf{w}[n-1]} \\ J[n] &= e^2[n] \\ \frac{\partial}{\partial\mathbf{w}}J[n]_{|\mathbf{w}=\mathbf{w}[n-1]} &= 2e[n]\frac{\partial}{\partial\mathbf{w}}e[n]_{|\mathbf{w}=\mathbf{w}[n-1]} = -2e[n]\mathbf{x}[n] \\ e[n] &= d[n] - \mathbf{w}^T[n-1]\mathbf{x}[n] \\ \Rightarrow \mathbf{w}[n] &= \mathbf{w}[n-1] + 2\mu e[n]\mathbf{x}[n] \end{aligned}$$

Optional:

$$\begin{aligned} e_f[n] &= x[n] + \Gamma[n]x[n-1] \\ \Gamma[n] &= \Gamma[n-1] - \mu\frac{\partial}{\partial\Gamma}J[n]|_{\Gamma=\Gamma[n-1]} \\ e_b[n] &= x[n-1] + \Gamma[n]x[n] \\ \frac{\partial}{\partial\Gamma}J[n] &= 2e_f[n]\frac{\partial}{\partial\Gamma}e_f[n]| + 2e_b[n]\frac{\partial}{\partial\Gamma}e_b[n] \\ &= 2e_f[n]x[n-1] + 2e_b[n]x[n] \\ \Rightarrow & \quad \Gamma[n] = \Gamma[n-1] - 2\mu(e_f[n]x[n-1] + e_b[n]x[n]) \end{aligned} \tag{30}$$

Note that LMS is easy to apply on any filter structure.

5.2 Verify the condition on the adaptation gain of the LMS algorithm, μ, for the convergence of the mean parameter vector $E\{\mathbf{w}[n]\}$, that is

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

Solution ____

LMS:

$$\begin{aligned}
\mathbf{w}[n] &= \mathbf{w}[n-1] + 2\mu e[n]\mathbf{x}[n] \\
&= \mathbf{w}[n-1] + 2\mu\big(d[n] - \mathbf{x}^T[n]\mathbf{w}[n-1]\big)\mathbf{x}[n] \\
&= \mathbf{w}[n-1] + 2\mu\mathbf{x}[n]\big(d[n] - \mathbf{x}^T[n]\mathbf{w}[n-1]\big) \\
\Rightarrow \mathbf{w}[n] &= \big(\mathbf{I} - 2\mu\mathbf{x}[n]\mathbf{x}^T[n]\big)\mathbf{w}[n-1] + 2\mu\mathbf{x}[n]d[n] \\
\Rightarrow \mathrm{E}\{\mathbf{w}[n]\} &= \big(\mathbf{I} - 2\mu\mathbf{R}\big)\mathrm{E}\{\mathbf{w}[n-1]\} + 2\mu\mathbf{r}_{\mathbf{x}d}
\end{aligned}$$

Invoke the independence assumptions between $\mathbf{x}[n]$ and $\mathbf{w}[n-1]$. Change coordinates and rotate (see notes) to obtain the *modes of convergence*

$$\begin{aligned}
\mathrm{E}\{\mathbf{v}[n]\} &= \big(\mathbf{I} - 2\mu\mathbf{\Lambda}\big)\mathrm{E}\{\mathbf{v}'[n-1]\} \\
\mathrm{E}\{v_j[n]\} &= (1 - 2\mu\lambda_j)^n\mathrm{E}\{v_j[0]\}
\end{aligned}$$

for every element of the weight error vector $v_j[n]$, $j = 1....N$ for the convergence of the mean

$$|(1 - 2\mu\lambda_j| < 1$$
$$\Rightarrow 0 < \mu < \frac{1}{\lambda_j}$$

5.3 Explain why the LMS algorithm is dependent on the spread of the power spectrum of the input of the adaptive filter, which is

$$\frac{P_{xx}^{max}(f)}{P_{xx}^{min}(f)}$$

How can preprocessing be used to overcome this dependence?

Solution

λ_{max} controls convergence limit, but λ_{min} corresponds to largest time constant.
$\lambda_{max} = \lambda_{min}$ is the best case, i.e. white input.

$$1 \leq \frac{\lambda_{max}}{\lambda_{min}} \leq \frac{P_{xx}^{max}(f)}{P_{xx}^{min}(f)}$$

Hence, the signal with large dynamic range will be inappropriate for LMS. Preprocessing, such as sub-optimal filtering, would act to reduce the spread of $\frac{\lambda_{max}}{\lambda_{min}}$.

5.4 Given the first order Markov (AR(1) process described by

$$x[n] = (1-\beta)w[n] + \beta x[n-1]$$

where $\beta \in (0,1)$ and $w[n]$ is a sequence of zero mean, unit power, IID noise; show its steady state power is

$$P = \frac{1-\beta}{1+\beta}$$

and that

$$\gamma = \frac{E\{x[n]x[n-1]\}}{E\{x^2[n]\}} = \beta.$$

Given the first order predictor

$$\hat{x}[n] = h\,x[n-1]$$

where h is the filter coefficient, evaluate the MSE as a function of h, the optimal predictor coefficient h_{opt}, the minimal error J_{min} and the corresponding eigenvalue. The prediction error is given by

$$e[n] = x[n] - h\,x[n-1] \tag{31}$$

Show that the $e[n] = x[n] - h_{opt}x[n-1]$ is independent of $x[n-1]$. Is this independence valid?

Solution

$$x^2[n] = (1-\beta)^2 w^2[n] + 2\beta w[n]x[n-1] + \beta^2 x^2[n-1]$$

$w[n]$ is independent of $x[n-1]$, which is built up with the past $w[n-j]$. Since $E\{w^2[n]\} = 1$

$$\begin{aligned} E\{x^2[n]\} &= (1-\beta)^2 + \beta^2 E\{x^2[n-1]\} \\ (1-\beta)^2 P &= (1-\beta)^2 \\ \Rightarrow P = \frac{1-\beta}{1+\beta} & \end{aligned}$$

$$\begin{aligned} x[n]x[n-1] &= (1-\beta)w[n]w[n-1] + \beta x^2[n-1] \\ \Rightarrow \Gamma[1] &= \frac{\beta P}{P} = \beta \end{aligned}$$

$$\begin{aligned} E(h) &= E(x^2[n]) - 2hE(x[n]x[n-1]) + h^2 E\{x^2[n-1]\} \\ &= P(1 - 2h\beta + h^2) \\ \Rightarrow h_{opt} &= \beta \end{aligned}$$

$$\begin{aligned} E_{min} &= E(h_{opt}) = P(1-\beta^2) = (1-\beta)^2 \\ \lambda &= P(1 \pm \beta) \\ \lambda_{max} &= P(1+\beta) = 1-\beta \end{aligned}$$

5.5 Derive the Recursive Least Squares (RLS) algorithm with forgetting factor λ, based on the sequential solution to

$$J[n] = \sum_{k=1}^{n} \lambda^{n-k} e^2[k] = \sum_{k=1}^{n} \lambda^{n-k} \mathbf{w}^T[n]\mathbf{x}[\mathbf{k}].$$

What is the role of λ?

Solution

For λ a forgetting factor, we have $0 << \lambda \leq 1$. Define the matrices

$$\begin{aligned} \mathbf{R}[n] &= \sum_{i=1}^{N} \lambda^{N-i} \mathbf{x}[i]\mathbf{x}^T[i] \\ \mathbf{Q}[n] &= \sum_{i=1}^{N} \lambda^{N-i} \mathbf{x}[i]d[i] \end{aligned}$$

$$\begin{aligned} \mathbf{R}[n] &= \lambda R[n-1] + \mathbf{x}[n]\mathbf{x}^T[n] \\ \mathbf{Q}[n] &= \lambda Q[n-1] + \mathbf{x}[n]d[n] \end{aligned} \tag{32}$$

Use the matrix inversion lemma as in notes:

$$\begin{aligned}
\mathbf{P}[n] &= \mathbf{R}^{-1}[n] \\
\mathbf{k}[n] &= \frac{\lambda^{-1}\mathbf{P}[n-1]\mathbf{X}[n]}{1+\lambda^{-1}\mathbf{X}[n]\mathbf{P}[n-1]\mathbf{X}[n]} \\
&= \lambda^{-1}[\mathbf{P}[n-1]-\mathbf{k}[n]\mathbf{X}^T[n]\mathbf{P}[n-1]]\mathbf{X}[n] \\
\mathbf{w}[n] &= \mathbf{w}[n-1]-\mathbf{k}[n]e[n] \\
e[n] &= d[n]-\mathbf{w}^T[n-1]\mathbf{X}[n]
\end{aligned}$$

5.6 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

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

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

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

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

(a) Let

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

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

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

be the *n*th-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 of $\mathbf{R}_x^{-1}$,

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

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

5.7 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

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

(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 λ_{max} and λ_{min} are bounded by the power spectrum as follows,

$$\begin{aligned}
\lambda_{\text{max}} &\leq \max_w \big[P_x(e^{jw})\big] = c\frac{1+\alpha}{1-\alpha} \\
\lambda_{\text{min}} &\geq \min_w \big[P_x(e^{jw})\big] = c\frac{1-\alpha}{1+\alpha}
\end{aligned}$$

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

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

then, as α increase, 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.

5.8 In recent years, there has been an increasing interest in nonlinear digital filters. This interest has included the design of adaptive nonlinear filters. Voltera systems are an important class of nonlinear filters. Assuming that $x(n)$ is real-valued, a second-order Volterra digital filter has the form

$$y(n) = \sum_{k=0}^{K} a(k)x(n-k) + \sum_{k_1=0}^{K_1} \sum_{k_2=k_1}^{K_2} b(k_1,k_2)x(n-k1)x(n-k2)$$

Note that the output, $y(n)$, is formed from a linear combination of first-order (linear) terms $x(n-k)$, and a linear combination of second-order (nonlinear) terms $x(n-k_1)x(n-k_2)$. As a specific example, consider the following second-order digital Volterra filter with time-varying coefficients,

$$y(n) = a_n(0)x(n) + a_n(1)x(n-1) + b_n(0)x^2(n) + b_n(1)x(n)x(n-1)$$

Let $\mathbf{\Theta}_n$ be the coefficient vector

$$\mathbf{\Theta}_n = \begin{bmatrix} a_n(0), & a_n(1), & b_n(0), & b_n(1) \end{bmatrix}^T$$

and let $\mathbf{x}(n)$ be the data vector

$$\mathbf{x}(n) = \begin{bmatrix} x(n), & x(n-1), & x^2(n), & x(n)x(n-1) \end{bmatrix}^T$$

(a) Using the LMS update equation

$$\mathbf{\Theta}_{n+1} = \mathbf{\Theta}_n - \frac{1}{2}\mu\nabla e^2(n)$$

where $e(n) = d(n) - y(n)$, derive the coefficient update equations for $a_n(0), a_n(1), b_n(0)$, and $b_n(1)$.

(b) What condition must be placed on μ in order for the coefficient vector $\mathbf{\Theta}_n$ to converge in the mean?

(c) Describe what happens if the third-order statistics of $x(n)$ are zero, i.e.,

$$\begin{aligned} E\{x^3(n)\} &= 0 \\ E\{x^2(n)x(n-1)\} &= 0 \\ E\{x(n)x^2(n-1)\} &= 0 \end{aligned}$$

Discuss how you might improve the performance of adaptive Volterra filter by having two step size parameters, μ_1 and μ_2, one for the linear terms and one for the nonlinear terms, and discuss how these parameters must be restricted in order for the filter to converge in the mean.

Solution

(a) The update equations are

$$\begin{aligned} a_{n+1}(0) &= a_n(0) + \mu e(n)x(n) \\ a_{n+1}(1) &= a_n(1) + \mu e(n)x(n-1) \\ b_{n+1}(0) &= b_n(0) + \mu e(n)x^2(n) \\ b_{n+1}(1) &= b_n(1) + \mu e(n)x(n)x(n-1) \end{aligned}$$

where

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

(b) The step size must satisfy the bound

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

where $\lambda_{\max}$ is the maximum eigenvalue of the autocorrelation matrix

$$\mathbf{R}_x = E\left\{ \begin{bmatrix} x(n) \\ x(n-1) \\ x^2(n) \\ x(n)x(n-1) \end{bmatrix} \begin{bmatrix} x(n) & x(n-1) & x^2(n) & x(n)x(n-1) \end{bmatrix} \right\}$$

(c) In this case, the autocorrelation matrix $\mathbf{R}_x$ defined above has the form

$$\mathbf{R}_x = \begin{bmatrix} E\{x^2(n)\} & E\{x(n)x(n-1)\} & 0 & 0 \\ E\{x(n)x(n-1)\} & E\{x^2(n-1)\} & 0 & 0 \\ 0 & 0 & E\{x^4(n)\} & E\{x^3(n)x(n-1)\} \\ 0 & 0 & E\{x^3(n)x(n-1)\} & E\{x^2(n)x^2(n-1)\} \end{bmatrix}$$

Since there is a decoupling of the adaptive filter coefficients, we may use *two* step sizes. One for the first two coefficients, and the second for the last two coefficients.

Notes:

Notes: