ANSWER <u>TWO QUESTIONS</u> FROM SECTION A AND <u>ONE QUESTION</u> FROM SECTION B. ALL QUESTIONS CARRY EQUAL MARKS.

<u>SECTION A</u>

Q1.

(a) Consider a causal linear time-invariant (LTI) system with the following transfer function relating its input $x(n)$ and output $y(n)$:

$$H(z) = \frac{2 - 2.4z^{-1} - 0.4z^{-2}}{(1 - 0.3z^{-1} - 0.4z^{-2})}.$$

i) Determine the linear constant coefficient difference equation for implementing the output $y(n)$.

(3 marks)

ii) Determine the poles of $H(z)$. **(2 marks)**

iii) Determine the impulse response, $h(n)$, (i.e. the response of the system to the unit impulse sequence $\delta(n)$) of this system.

(6 marks)

iv) Is the system stable? If the above system is now non-causal and has the same transfer function $H(z)$, is it still stable? Explain.

(4 marks)

v) Determine the z-transform of $x(n) = e^{j(n\omega_0)}u(n)$, where $u(n)$ is the unit step sequence.

(2 marks)

vi) Determine the output $y(n)$ of the system to $x(n) = e^{j(n\omega_0)}u(n)$. Identify the transient and steady state responses in your solution.

(6 marks)

(b)

i) Determine the impulse response $h(n)$ of a highpass filter with the following ideal discrete-time frequency response:

$$H_{hp}(e^{j\omega}) = \begin{cases} 0 & 0 \le |\omega| < \omega_c \\ e^{-j\omega M/2} & \omega_c \le |\omega| \le \pi. \end{cases}$$ **(5 marks)**

[The DTFT of the sequence $x(n)$ is $X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-jn\omega}$.

The inverse discrete time Fourier transform of a function $X(e^{j\omega})$ is

$$x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\omega})e^{jn\omega}d\omega.]$$

ii) The input to a **causal linear time-invariant (LTI)** system with transfer function $H(z)$ is

$$x(n) = u(-n-1) + (1/2)^n u(n)$$

with z-transform $X(z) = \dfrac{-1}{1-z^{-1}} + \dfrac{1}{1-\frac{1}{2}z^{-1}}, \quad \dfrac{1}{2} < |z| < 1.$

Determine $y(n)$, if the z-transform of the system output $y(n)$ is

$$Y(z) = \frac{-\frac{1}{2}z^{-1}}{(1-\frac{1}{2}z^{-1})(1+z^{-1})}.$$ **(5 marks)**

[Hint: Note, one of the poles of $X(z)$, which limits the ROC of $X(z)$, is cancelled by the zero of $H(z)$, which is a causal LTI system. Therefore, the ROC of $Y(z)$ is the region in the z-plane that satisfies the remaining two constraints.]

Q2.

(a) Consider the discrete-time processing system in Figure Q2-1. It is required to perform an equivalent filtering in the continuous time domain with the following specifications:

1. The gain $|H_{eff}(j\Omega)|$ should be within ± 0.01 of unity (zero dB) in the frequency band $0 \le \Omega \le 2\pi(2000)\ Hz$.
2. The gain should be no greater than ± 0.001 in the frequency band $2\pi(2500)\ Hz \le \Omega$.

Ω is the continuous time radian frequency in radians.

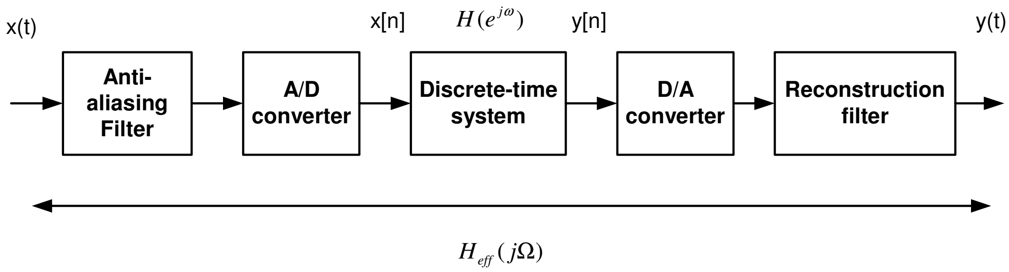


Figure Q2-1.

i) If the sampling rate is 10^4 samples per second, determine the equivalent specifications of the discrete-time system.

(4 marks)

ii) Using the following formulae, estimate the parameters M and β of the Kaiser window required to satisfy the specifications.

$$\beta = \begin{cases} 0.1102(A-8.7) & A > 50 \\ 0.5842(A-21)^{0.4} + 0.07886(A-21), & 21 \le A \le 50; \\ 0.0 & 0 \le A < 21 \end{cases}$$

$$M = (A-8)/(2.285\cdot\Delta\omega),$$

where $A = -20\log_{10}\delta$ (dB), $\Delta\omega = |\omega_s - \omega_p|$, δ is the minimum amplitude of the passband and stopband ripples, ω_s is the stopband cutoff frequency, and ω_p is the passband cutoff frequency. **(4 marks)**

iii) Use the following formula to estimate the order of the linear-phase filter if the McClellan-Parks design algorithm is used.

$$N = \frac{-20\log_{10}\sqrt{\delta_1\delta_2} - 13}{14.6\Delta f},$$

where δ_1, δ_2 are respectively the maximum amplitude of the passband and stopband ripples and $\Delta f = \Delta\omega / 2\pi$. Comment on the two algorithms.

(5 marks)

(c) It is required to design a circular symmetric two dimensional lowpass filter by transforming a 1-D zero-phase lowpass prototype $h(n)$ with frequency response $H(\omega)$.

Suppose that the 1D prototype can be written as:

$$\begin{aligned} H(\omega) &= h(0) + \sum_{n=1}^{N} h(n)[\exp(-j\omega n) + \exp(j\omega n)] \\ &= \sum_{n=0}^{N} a(n)\cos(n\omega) = \sum_{n=0}^{N} a(n) T_n[\cos(\omega)], \end{aligned} \qquad (*)$$

where $a(n) = \begin{cases} h(0), n = 0 \\ 2h(n), n > 0 \end{cases}$ and $T_n(x)$ is the *n-th* order Chebyshev polynomial. The 2D frequency response is obtained by substituting $\cos\omega$ in equation (*) by $F(\omega_1, \omega_2)$ to obtain:

$$H(\omega_1, \omega_2) = \sum_{n=0}^{N} a(n) T_n[F(\omega_1, \omega_2)].$$

Let the transformation function be

$$F(\omega_1, \omega_2) = \tfrac{1}{2}(-1 + \cos\omega_1 + \cos\omega_2 + \cos\omega_1\cos\omega_2).$$

i) Show that $H(\pm\omega_2, \pm\omega_1) = H(\omega_1, \omega_2)$. **(3 marks)**

ii) Using Figure Q2-2, indicate the passband and stopband of $H(\omega_1, \omega_2)$ if the passband and stopband cutoff frequencies of $H(\omega)$ are respectively 0.4π and 0.5π. Comment on the circular symmetry of the frequency response if the cutoff frequencies are increased to say 0.8π and 0.9π.

(4 marks)

iii) Briefly compare the least squares and the McClellan transformation methods for designing two-dimensional FIR filters in terms of i) the spectral support of the digital filters to be designed, and ii) the peak approximation error.

(2 marks)

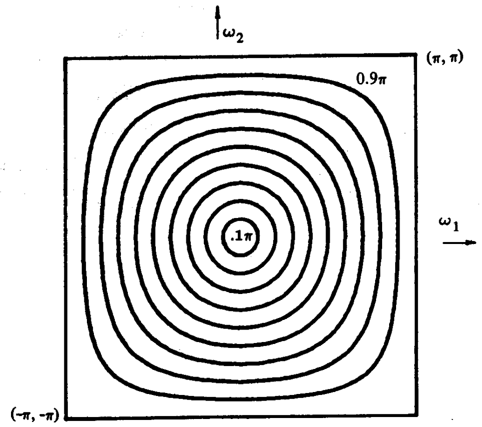


Figure Q2-2. Contours of constant value of $F(\omega_1, \omega_2)$. The contour marked with 0.9π means $F(\omega_1, \omega_2) = \cos(0.9\pi)$ and vice versa.

(c)

It is desired to design a norm constrained 2D FIR filter by second order cone programming (SOCP) method, which can be written as:

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}} & \boldsymbol{c}^T \boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i\|_2 \le \boldsymbol{c}_i^T \boldsymbol{x} + d_i, \;\; i = 1, \ldots, M, \\ & \boldsymbol{F}\boldsymbol{x} = \boldsymbol{g}, \end{array}$$

where $\boldsymbol{x} \in \Re^N$ is the optimization variable, $\boldsymbol{A}_i \in \Re^{N_i \times N}$, $\boldsymbol{F} \in \Re^{P \times N}$, $\boldsymbol{b}_i \in \Re^{N_i}$, $\boldsymbol{c}, \boldsymbol{c}_i \in \Re^N$, $\boldsymbol{g} \in \Re^P$, $d_i \in \Re$ and $\|\boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i\|_2 \le \boldsymbol{c}_i^T \boldsymbol{x} + d_i$ for $i = 1, \ldots, M$ are second-order cone constraints.

Consider a 2D FIR filter with frequency response: $H(e^{j\omega_1}, e^{j\omega_2}) = \sum\limits_{n_1=0}^{N_1-1} \sum\limits_{n_2=0}^{N_2-1} h(n_1, n_2) e^{-jn_1\omega_1} e^{-jn_2\omega_2}$.

i) Show that $H(e^{j\omega_1}, e^{j\omega_2}) = \boldsymbol{h}^T(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega}))$

where $\boldsymbol{\omega} = [\omega_1, \omega_2]^T$, $\boldsymbol{h} = [h(0,0), h(0,1), \ldots, h(N_1 - 1, N_2 - 2), h(N_1 - 1, N_2 - 1)]^T$, $\boldsymbol{c}(\boldsymbol{\omega}) = [1, \cos(\omega_2), \ldots, \cos((N_1 - 1)\omega_1 + (N_2 - 1)\omega_2)]^T$ and

$\boldsymbol{s}(\boldsymbol{\omega}) = [0, \ldots, \sin((N_1 - 1)\omega_1 + (N_2 - 1)\omega_2)]^T$.

(2 marks)

ii) Let the desired response in the passband $\omega_1, \omega_2 \in \Omega_p$ and stopband $\omega_1, \omega_2 \in \Omega_s$ be $H_d(e^{j\omega_1}, e^{j\omega_2})$. It is desired to design the FIR filter using the weighted least squares criterion subject to a set of peak magnitude error constraints around the band edges $\omega_1, \omega_2 \in \Omega_E$. The corresponding filter design problem can be written as:

$$\min_{\boldsymbol{h}} \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega}) \,|\, H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}) \,|^2 d\boldsymbol{\omega},$$

subject to $|\, H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}) \,| \le \delta_p, \ \omega_1, \omega_2 \in \Omega_E$,

where $W(\boldsymbol{\omega})$ is a real positive weighting function and δ_p is a prescribed peak ripple. Show that the WLS function is given by

$$\begin{aligned}\mathrm{WLS}(\boldsymbol{h}) &= \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega}) \,|\, H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}) \,|^2 d\boldsymbol{\omega} \\ &= \boldsymbol{h}^T \boldsymbol{Q} \boldsymbol{h} - 2\boldsymbol{h}^T \boldsymbol{g} + c,\end{aligned}$$

where $\boldsymbol{Q} = \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega})(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega}))(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega}))^H d\boldsymbol{\omega}$,

$$\boldsymbol{g} = \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega}) \operatorname{Re}\{(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega})) H_d^*(e^{j\omega_1}, e^{j\omega_2})\} d\boldsymbol{\omega},$$

$c = \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega}) \,|\, H_d(e^{j\omega_1}, e^{j\omega_2}) \,|^2 \, d\boldsymbol{\omega}$ and superscript ^{H} stands for conjugate transpose.

[Hint:

$$\begin{aligned}&|\, H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}) \,|^2 \\ &\quad = (H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}))(H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}))^*]\end{aligned}$$

(3 marks)

iii) In norm constrained digital filter design, the 2-norm of the filter coefficients, $\|\boldsymbol{h}\|_2^2 = \boldsymbol{h}^T \boldsymbol{h}$, is also minimized. The objective function to be minimized is thus

$$E(\boldsymbol{h}) = \mathrm{WLS}(\boldsymbol{h}) + \lambda \|\boldsymbol{h}\|_2^2,$$

where λ is a constant to control the importance of the two terms above.

By discretizing ω_1, ω_2 uniformly in $\omega_1, \omega_2 \in \Omega_E$ into M points, $\boldsymbol{\omega}_i$ for i=1,…,M, show that the above problem can be written as:

$$\begin{aligned}&\min_{\boldsymbol{h},\delta} \quad \delta \\ &\text{subject to} \quad \delta_p - \{\alpha_R^2(\boldsymbol{\omega}_i) + \alpha_I^2(\boldsymbol{\omega}_i)\}^{1/2} \ge 0, \ \ i = 1, \ldots, M, \\ &\qquad \|\tilde{\boldsymbol{L}}^T \boldsymbol{h} - \tilde{\boldsymbol{L}}^{-1} \boldsymbol{g}\|_2 \le \delta \\ &\alpha_R(\boldsymbol{\omega}) = W(\boldsymbol{\omega}) \cdot \{\boldsymbol{h}^T \boldsymbol{c}(\boldsymbol{\omega}) - \operatorname{Re}[H_d(e^{j\omega_1}, e^{j\omega_2})]\},\end{aligned}$$

$$\alpha_I(\boldsymbol{\omega}) = W(\boldsymbol{\omega}) \cdot \{\boldsymbol{h}^T \boldsymbol{s}(\boldsymbol{\omega}) + \mathrm{Im}[H_d(e^{j\omega_1}, e^{j\omega_2})]\}.$$

Express $\tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^T$ in terms of $\boldsymbol{Q}$. **(6 marks)**

[Hint: If $\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^T$ where $\tilde{\boldsymbol{L}}$ is some nonsingular upper triangular matrix, then

$$\min_{\boldsymbol{h}} \quad \boldsymbol{h}^T \tilde{\boldsymbol{Q}} \boldsymbol{h} - 2\boldsymbol{h}^T \boldsymbol{g} + c$$

is equivalent to $\quad \min_{\boldsymbol{h}} \| \tilde{\boldsymbol{L}}^T \boldsymbol{h} - \tilde{\boldsymbol{L}}^{-1} \boldsymbol{g} \|_2^2$,

where the $L2$ norm of a vector $\boldsymbol{x}$ is given by $\| \boldsymbol{x} \|_2^2 = \boldsymbol{x}^T \boldsymbol{x}$. Find the new $\tilde{\boldsymbol{Q}}$ of $E(\boldsymbol{h})$ using Part (ii).]

Q3.

a) Let $x(n)$ be a wide-sense stationary random process with mean μ_x and autocorrelation $R_{xx}(k)$. Let $y(n)$ be the output obtained by applying $x(n)$ as input to a stable linear time-invariant system with impulse response $h(n)$ and transfer function $H(e^{j\omega})$.

i) Show that $y(n)$ is also a wide-sense stationary process with mean μ_y and autocorrelation $R_{yy}(k)$ given by

$$\mu_y = H(1) \cdot \mu_x,$$

and $\quad R_{yy}(k) = R_{xx}(k) * h(k) * h(-k)$,

where * denotes the discrete-time convolution operation.

(4 marks)

ii) Hence, show that the Power Spectral Density (PSD) of $y(n)$ is

$$S_{yy}(e^{j\omega}) = S_{xx}(e^{j\omega}) \left| H(e^{j\omega}) \right|^2.$$ **(3 marks)**

b) Figure Q3-1 shows the signal flow graph of a causal stable second-order direct form II system with system function $H(z)$ and impulse response $h(n)$. Suppose that the results of the multiplications $\{ a_1 \cdot w(n-1),\ a_2 \cdot w(n-2) \}$, and $\{ b_0 \cdot w(n),\ b_1 \cdot w(n-1),\ b_2 \cdot w(n-1) \}$ are rounded to $B_a + 1$ and $B_b + 1$ bits, respectively and let $e_{a_i}(n)$, i=1,2, and $e_{b_k}(n)$, k=0,1,2, be the corresponding quantization noises. Let $\xi_{a_i}(n)$, i=1,2, be the quantization noises at the system output generated respectively by $e_{a_i}(n)$, i=1,2. Assume that:

1) $e_{a_i}(n)$ and $e_{b_k}(n)$ are zero-mean wide-sense stationary white-noises, and are uncorrelated with each other.
2) Each noise source has a uniform distribution of amplitudes over one quantization interval.
3) $e_{a_i}(n)$ and $e_{b_k}(n)$, are uncorrelated with the input of the corresponding quantizers, and the input to the system.

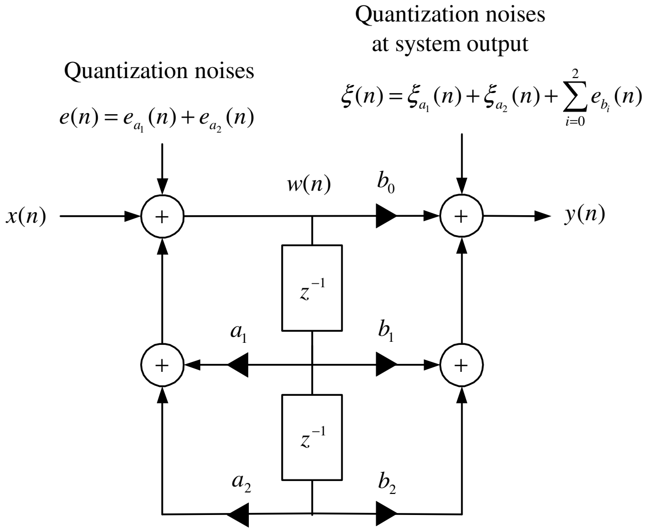


Figure Q3-1.

i) Determine the transfer function $H(z) = Y(z)/X(z)$. **(3 marks)**

ii) Show that

$$E[\{\xi_{a_1}(n) + \xi_{a_2}(n) + \sum_{i=0}^{2} e_{b_i}(n)\}^2] = E[\xi_{a_1}^2(n)] + E[\xi_{a_2}^2(n)] + \sum_{i=0}^{2} E[e_{b_i}^2(n)].$$

(5 marks)

iii) If the variances of $e_{a_i}(n)$ and $e_{b_j}(n)$ are given by $\sigma_{e_{a_i}}^2 = \dfrac{2^{-2B_a}}{12}$ and $\sigma_{e_{b_i}}^2 = \dfrac{2^{-2B_b}}{12}$ respectively, determine the power spectral densities, $P_{a_i}(\omega)$ and $P_{b_j}(\omega)$, of $e_{a_i}(n)$ and $e_{b_j}(n)$.

[Hint: $\sigma_x^2 = \dfrac{1}{2\pi}\int_{-\pi}^{\pi} P_x(\omega)d\omega$, where σ_x^2 and $P_x(\omega)$ are respectively the variance and power spectral density of a random process $x(n)$.] **(3 marks)**

iv) Show that the total noise variance at the output is

$$\sigma_\xi^2 = 3\frac{2^{-2B_b}}{12} + 2\frac{2^{-2B_a}}{12}\sum_{n=-\infty}^{\infty} |h(n)|^2 .$$

[Hint: You might use the Parseval's theorem $\sum_{n=-\infty}^{\infty} |h(n)|^2 = \dfrac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega$,

or evaluating the result in b(ii) directly.] **(5 marks)**

c) Consider a zero mean real random vector $\boldsymbol{x} = [x(1),..., x(M)]^T$ with covariance matrix $\boldsymbol{R}_{xx} = E[\boldsymbol{x}\boldsymbol{x}^T]$. The total variance of $\boldsymbol{x}$ is $\sigma_x^2 = E[\boldsymbol{x}^T\boldsymbol{x}] = \sum_{i=1}^{M} E[x^2(i)] = \sum_{i=1}^{M} \sigma_{x(i)}^2$. Consider the transformed vector $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} = [y(1),..., y(M)]^T$, where $\boldsymbol{C}$ is an orthogonal transformation.

i) Show that the total variance of the transformation vector $\boldsymbol{y}$ $\sigma_y^2 = E[\boldsymbol{y}^T\boldsymbol{y}] = \sum_{i=1}^{M} \sigma_{y(k)}^2$ is also σ_x^2. That is, the total variance of a random vector before and after an orthogonal transformation remains unchanged. **(2 marks)**
[Hint: For orthogonal matrix $\boldsymbol{C}$, $\boldsymbol{C}^T\boldsymbol{C} = \boldsymbol{I}$.]

ii) Suppose that the k-th transformed coefficient, $y(k)$, is to be quantized by a uniform quantizer with b_k number of bits. The total number of bits, b, for representing each realization of the random vector $\boldsymbol{y}$ is then given by $b = \sum_{k=1}^{M} b_k$.
Let the quantization error be $q(k) = y(k) - \hat{y}(k)$ where $\hat{y}(k)$ is the corresponding quantized value of the transformed coefficient. Similarly, let the corresponding error in the original domain be $\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{C}^T\hat{\boldsymbol{y}}$ where $\boldsymbol{e} = [e(1),..., e(M)]^T$ and $\hat{\boldsymbol{y}} = [\hat{y}(1),..., \hat{y}(M)]^T$.

Show that $\sigma_e^2 = \sum_{k=1}^{M} \sigma_{q(k)}^2 = \sum_{k=1}^{M} E[q^2(k)]$. **(3 marks)**
[Hint: use the results in (i) and the fact that $\boldsymbol{x} = \boldsymbol{C}^T\boldsymbol{y}$ and $\boldsymbol{q} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ with $\boldsymbol{q} = [q(1),..., q(M)]^T$.]

iii) Assume that all the quantizers can be modeled by the following formula:

$$\sigma_{q(k)}^2 = c \cdot 2^{-2b_k} \sigma_{y(k)}^2 \qquad (*)$$

where c is a constant.
Show that $\sigma_e^2 \geq cM \cdot 2^{-\frac{2}{M}b} \cdot \left(\prod_{k=1}^{M} \sigma_{y(k)}^2\right)^{1/M}$. **(3 marks)**

[Hint: Let $a_k = 2^{-2b_k} \sigma_{y(k)}^2 \geq 0$ and substitute (*) into the result of part (ii). Then, use the fact that arithmetic mean is greater or equal to the geometric mean for nonnegative number a_k, i.e. $\frac{1}{M}\sum_{k=1}^{M} a_k \geq \left(\prod_{k=1}^{M} a_k\right)^{1/M}$, with equality if and only if $a_1 = ... = a_M$.]

iv) Since the right hand side of the inequality in part (iii) depends only on b and $\left(\prod_{k=1}^{M} \sigma_{y(k)}^2\right)^{1/M}$, but not on b_k, it can be made equal by choosing b_k such that the average distortion is minimized with $a_1 = ... = a_M$. This requires $a_k = \sigma^2_{q(k)}$, k=1,…,M, to be identical. Let's denote the corresponding minimum average distortion in time domain by $\sigma^2_{e,T\min}$. If PCM is employed to quantize the components of vector $\boldsymbol{x}$ directly, the number of bits per coefficient used will be b/M. The corresponding average distortion of $x(k)$ is $\sigma^2_{x(k)} = c \cdot 2^{-2b/M} \sigma^2_{x(k)}$. Denote the corresponding total average distortion by $\sigma^2_{e,PCM} = \sum_{k=1}^{M} \sigma^2_{x(k)}$. Using the results in parts (i) to (iii), show that the coding gain of using transformation over PCM is given by

$$\frac{\sigma^2_{e,PCM}}{\sigma^2_{e,T\min}} = \frac{\frac{1}{M}\sum_{k=1}^{M}\sigma^2_{y(k)}}{\left(\prod_{k=1}^{M}\sigma^2_{y(k)}\right)^{1/M}}.$$

(2 marks)

Q4.

a) It is required to design an FIR Wiener filter $\{w_i, i = 0,..., L-1\}$ as shown in Figure Q4-1 to approximate a desired signal $d(n)$ from its input signal $x(n)$. The Wiener filter is of L taps and its output $y(n)$ is given by:

$$y(n) = \sum_{k=0}^{L-1} w_k x(n-k) = \boldsymbol{W}^T \boldsymbol{X}_n,$$

where $\boldsymbol{W} = \begin{bmatrix} w_0 & w_1 & . & . & w_{L-1} \end{bmatrix}^T$ is the weight vector, and
$\boldsymbol{X}_n = \begin{bmatrix} x(n) & x(n-1) & . & . & x(n-L+1) \end{bmatrix}^T$ is the input signal vector.

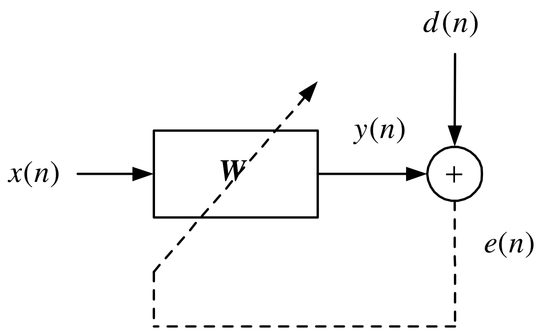


Figure Q4-1.

Assume that all the signals are *stationary*.

i) Let $e(n) = d(n) - y(n)$ be the approximation error. Show that the Mean Squared Error (MSE)

$$\xi = E[e^2(n)]$$

can be written as:

$$\xi = r_{dd}(0) + \boldsymbol{W}^T \boldsymbol{R}\boldsymbol{W} - 2\boldsymbol{P}^T \boldsymbol{W} ,$$

where $r_{dd}(0) = E[d^2(n)]$,

$\boldsymbol{R} = E[\boldsymbol{X}_n \boldsymbol{X}_n^T] = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \ldots & r_{xx}(L-1) \\ r_{xx}(1) & r_{xx}(0) & \ldots & r_{xx}(L-2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{xx}(L-1) & r_{xx}(L-2) & \ldots & r_{xx}(0) \end{bmatrix}$ is the autocorrelation matrix of the input signal vector, and

$\boldsymbol{P} = E[d(n)\boldsymbol{X}_n] = \begin{bmatrix} r_{xd}(0) & r_{xd}(1) & \ldots & r_{xd}(L-1) \end{bmatrix}^T$ is the cross correlation vector, and E and T denote the expectation operator and matrix transposition, respectively.

$r_{xd}(n)$ denotes the cross-correlation between signals $x(n)$ and $d(n)$ and $r_{xx}(n)$ denotes the autocorrelation of signal $x(n)$.

(4 marks)

ii) Using the result in part (i), show that the optimum weight vector $\boldsymbol{W}^*$ that minimizes the MSE is

$$\boldsymbol{W}^* = \boldsymbol{R}^{-1}\boldsymbol{P} .$$ **(3 marks)**

iii) In recursive least squares (RLS) adaptive filtering algorithm, the autocorrelation matrix $\boldsymbol{R}(n)$ and cross correlation vector $\boldsymbol{P}(n)$ at the *n-th* time instant are respectively estimated recursively as:

$$\boldsymbol{R}(n) = \lambda \boldsymbol{R}(n-1) + \boldsymbol{X}_n \boldsymbol{X}_n^T \text{ and } \boldsymbol{P}(n) = \lambda \boldsymbol{P}(n-1) + d(n)\boldsymbol{X}_n ,$$

where $0 < \lambda < 1$ is a forgetting factor, $\boldsymbol{R}(0) = \delta \boldsymbol{I}$ and $\boldsymbol{P}(0) = \boldsymbol{0}$ with δ a small positive number.

Show that $\boldsymbol{R}(n) = \lambda^n \delta \boldsymbol{I} + \tilde{\boldsymbol{X}}_n^T \tilde{\boldsymbol{X}}_n$ and $\boldsymbol{P}(n) = \tilde{\boldsymbol{X}}_n^T \boldsymbol{d}_n$,

where $\tilde{\boldsymbol{X}}_n^T = [\lambda^{(n-1)/2}\boldsymbol{X}_1, \lambda^{(n-2)/2}\boldsymbol{X}_2, ..., \lambda^{1/2}\boldsymbol{X}_{n-1}, \boldsymbol{X}_n]$ and $\boldsymbol{d}_n = [d(1), ..., d(n)]^T$.

Show that $\boldsymbol{R}(n) = \boldsymbol{C}_n^T \boldsymbol{C}_n$ and $\boldsymbol{P}(n) = \boldsymbol{C}_n^T \tilde{\boldsymbol{d}}_n$, where $\boldsymbol{C}_n = \begin{bmatrix} \sqrt{\delta \lambda^n}\boldsymbol{I} \\ \tilde{\boldsymbol{X}}_n \end{bmatrix}$ and $\tilde{\boldsymbol{d}}_n = [\boldsymbol{0}_L^T, \boldsymbol{d}_n^T]^T$, where $\boldsymbol{0}_L$ is a zero column vector of size L.

Suppose $\boldsymbol{C}_n$ can be factored recursively using the QR decomposition so that $\boldsymbol{Q}_n^T[\boldsymbol{C}_n, \tilde{\boldsymbol{d}}_n] = [\boldsymbol{R}_n, \boldsymbol{d}_n^{'}]$, where, $\boldsymbol{Q}_n$ is some orthogonal matrix and $\boldsymbol{R}_n$ is an (LxL) upper triangular matrix. Assuming that $\boldsymbol{R}_n$ is nonsingular, show that the normal equation

$$\boldsymbol{R}(n)\boldsymbol{W}(n) = \boldsymbol{P}(n)$$

can be simplified to $\boldsymbol{R}_n\boldsymbol{W}(n) = \boldsymbol{d}_n^{'}$.

[Hint: for orthogonal matrix $\boldsymbol{Q}_n$, $\boldsymbol{Q}_n^T\boldsymbol{Q}_n = \boldsymbol{I}$] (**6 marks**)

b) In the least mean squares (LMS) algorithm, the weight vector is updated recursively as follows

$$\boldsymbol{W}_{n+1} = \boldsymbol{W}_n + 2\mu e(n)\boldsymbol{X}_n,$$

where $e(n) = d(n) - \boldsymbol{W}_n^T\boldsymbol{X}_n$ and μ is an appropriately chosen stepsize parameter to ensure convergence of the algorithm. Briefly explain the derivation of this equation. Suggest a method to determine the stepsize parameter without having to compute the eigenvalues of $\boldsymbol{R}$. Comment on the selection of the stepsize parameter in stationary and time-varying environment.

(6 marks)

c) Consider the linear predictor for a wide-sense stationary process $x(n)$ as shown in Figure Q4-2.

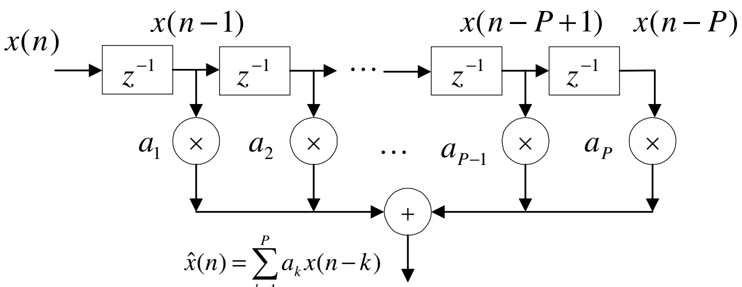


Figure. Q4-2 : *P-th* order Linear Predictor.

Let $e(n) = x(n) - \hat{x}(n)$ be the prediction error and MSE be the mean squared error given by:

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

i) Show that:

$$MSE = r_{xx}(0) - 2\boldsymbol{a}^T\boldsymbol{r}_{xx} + \boldsymbol{a}^T\boldsymbol{R}_{xx}\boldsymbol{a},$$

where $\boldsymbol{R}_{xx} = E[\boldsymbol{x}\boldsymbol{x}^T]$ is the *autocorrelation matrix* of the input signal,

$\boldsymbol{r}_{xx} = E[x(n)\boldsymbol{x}]$ is the *autocorrelation* vector of the input signals,

$$\boldsymbol{a}^T = [a_1, \quad a_2, \quad ..., \quad a_P] \text{ and } \boldsymbol{x}^T = [x(n-1) \quad x(n-2), \quad ..., \quad x(n-P)].$$

(5 marks)

ii) By differentiating *MSE* with respect to a_i, show that the optimal linear predictor coefficients with minimum *MSE* is given by

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & r_{xx}(2) & \cdots & r_{xx}(P-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(P-2) \\ r_{xx}(2) & r_{xx}(1) & & \cdots & r_{xx}(P-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}(P-1) & r_{xx}(P-2) & r_{xx}(P-3) & \cdots & r_{xx}(0) \end{bmatrix}^{-1} \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ \vdots \\ r_{xx}(P) \end{bmatrix}.$$

(3 marks)

iii) By taking the z-transform of the linear prediction equation

$$e(n) = x(n) - \sum_{i=1}^{P} a_i x(n-i),$$

show that the transfer function from $x(n)$ to the prediction error $e(n)$ is

$$\frac{E(z)}{X(z)} = H(z) = 1 - \sum_{i=1}^{P} a_i z^{-i},$$

where $E(z)$ is the z-transform of $e(n)$.

If $H(z) = 1 - \sum_{i=1}^{P} a_i z^{-i}$ has all its zeros ω_k except $\omega_k = 0, \pi$, $k=1,\ldots,P$, being distinct and are on the unit circle with $E(z) = e$ and ROC |z|>1, show that $x(n)$ can be written as a sum of P complex sinusoids as follows

$$x(n) = \sum_{k=1}^{P} A_k e^{j(n\omega_k)} u(n),$$

for some constant $A_k = |A_k| e^{j\phi_k}$. Moreover, if P is even and a_i's are real numbers, $x(n)$ can be written as

$$x(n) = [2\sum_{k=1}^{P/2} |A_k| \cos(\omega_k n + \phi_k)] \cdot u(n). \qquad (*)$$

(5 marks)

iv) Using the result in (iii) and assuming that the additive noise is small, suggest a method to estimate the frequencies of a multi-sinusoidal signal in form of (*) above.

(1 marks)

SECTION B

Q5

a) What is meant by an unbiased estimator and the bias of an estimator?

(4 marks)

b) State the mean square error criterion for measuring the performance of estimators.

(2 marks)

c) Explain the concept of minimum variance unbiased (MVU) estimator.

(2 marks)

d) By means of graphical illustration, explain why a MVU estimator may not exist.

(3 marks)

e) Consider the problem of nonlinear fitting with the following observations

$$x[n] = f(\boldsymbol{\beta}_n, \boldsymbol{\theta}) + w[n],\ n = 0,1,\ldots,N-1,$$

where $\boldsymbol{\beta}_n$, $n = 0,1,\ldots,N-1$, are known vectors, $\boldsymbol{\theta}$ is the parameter to be estimated, $f(\boldsymbol{\beta}_n, \boldsymbol{\theta})$ is a twice continuously differentiable function in $\boldsymbol{\theta}$, and $w[n]$ is a zero mean white Gaussian noise process with variance σ^2.

Given: the probability distribution function (PDF) of a Gaussian distribution with mean m and variance σ^2 is given by $p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp[-\frac{1}{2\sigma^2}(x-m)^2]$.

i. Write down the likelihood function $p(\boldsymbol{x};\boldsymbol{\theta})$ where $\boldsymbol{x} = (x[0],\ldots,x[N-1])^T$.

(3 marks)

ii. Determine $\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \theta_i}$ and $\frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$, i=1,2.

(6 marks)

iii. Show that the Fisher information matrix, $[\boldsymbol{I}(\boldsymbol{\theta})]_{ij} = -E_w\left[\frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right]$, is given

$$\boldsymbol{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1} \nabla_\theta f(\boldsymbol{\beta}_n;\boldsymbol{\theta})) \cdot \nabla_\theta f(\boldsymbol{\beta}_n;\boldsymbol{\theta})^T,$$

where $\nabla_\theta f(\boldsymbol{\beta}_n;\boldsymbol{\theta})$ is the gradient of $f(\boldsymbol{\beta}_n;\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

[Hint: You can assume that, $x[n] - f(\boldsymbol{\beta}_n;\boldsymbol{\theta}) = w[n]$, which is the value at the true parameter. This is reasonable since the estimator is assumed to be unbiased.]

For linear model, $f(\boldsymbol{\beta}_n;\boldsymbol{\theta}) = \boldsymbol{\beta}_n^T \boldsymbol{\theta}$, determine the corresponding Fisher information matrix.

(4 marks)

iv. Determine the Cramer-Rao Lower Bound (CRLB) of the following line fitting problems with observations

$$x[n] = A + Bn + w[n],\ n = 0,1,\ldots,N-1,$$

where $w[n]$ is a zero mean white Gaussian noise process with variance σ^2, and A and B are respectively the intercept and slope of the straight line $A + Bn$ to be estimated. The parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ is $[A, B]^T$.

[Hint: $\boldsymbol{C}_{\hat{\theta}_i} \geq [\boldsymbol{I}^{-1}(\boldsymbol{\theta})]_{ii}$. For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the matrix inverse is $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. $\sum_{n=0}^{N-1} n^2 = \frac{N(N-1)(2N-1)}{6}$.]

(6 marks)

v. Show that maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ for the nonlinear fitting problem satisfies:

$$\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n] - f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot \nabla f_{\boldsymbol{\theta}}(\boldsymbol{\beta}_n;\boldsymbol{\theta}) = 0.$$

For linear model, $f(\boldsymbol{\beta}_n;\boldsymbol{\theta}) = \boldsymbol{\beta}_n^T\boldsymbol{\theta}$, show that the corresponding MLE satisfies

$$\left(\sum_{n=0}^{N-1}\boldsymbol{\beta}_n\boldsymbol{\beta}_n^T\right)\boldsymbol{\theta} = \sum_{n=0}^{N-1} x[n]\boldsymbol{\beta}_n,$$

which is the normal equation equation in linear estimation.

(3 marks)

Q6

a) What is the main difference and possible advantage between the classical approach to statistical estimation such as maximum likelihood estimation and the Bayesian approach?

(4 marks)

b) The Bayesian mean square error (Bmse) is defined as

$$Bmse(\hat{\boldsymbol{\theta}}) = \int\int \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2^2\, p(\boldsymbol{x},\boldsymbol{\theta})d\boldsymbol{x}d\boldsymbol{\theta},$$

where the expectation is performed with respect to the joint PDF $p(\boldsymbol{x},\boldsymbol{\theta})$.

Using the Bayes' theorem $p(\boldsymbol{x},\boldsymbol{\theta}) = p(\boldsymbol{\theta}\mid\boldsymbol{x})p(\boldsymbol{x})$, show that the estimator which minimizes the Bmse is given by

$$\hat{\boldsymbol{\theta}} = \int \boldsymbol{\theta}\cdot p(\boldsymbol{\theta}\mid\boldsymbol{x})d\boldsymbol{\theta} = E_{\boldsymbol{\theta}}[\boldsymbol{\theta}\mid\boldsymbol{x}].$$

(5 marks)

c) Consider the following state space model over time instants k=0,1,…

(6-1) $\boldsymbol{\theta}_k = \boldsymbol{A}_k \boldsymbol{\theta}_{k-1} + \boldsymbol{\varepsilon}_k$,

(6-2) $\boldsymbol{x}_k = \boldsymbol{H}_k \boldsymbol{\theta}_k + \boldsymbol{w}_k$,

where $\boldsymbol{\theta}_k$ is the mx1 state vector, $\boldsymbol{A}_k$ is the mxm state transition matrix, $\boldsymbol{\varepsilon}_k$ is a zero mean Gaussian distributed noise (commonly referred to as excitation/innovation) with covariance $\boldsymbol{C}_{\varepsilon_k}$, $\boldsymbol{x}_k$ is the dx1 measurement vector, $\boldsymbol{H}_k$ is the dxm measurement matrix, and $\boldsymbol{w}_k$ is the dx1 measurement noise which is assumed to be Gaussian distributed with mean zero and covariance $\boldsymbol{C}_{w_k}$.

i) Show that the mean and covariance of $\boldsymbol{\theta}_k$ given $\boldsymbol{\theta}_{k-1}$ and the dynamical equation (6-1) satisfy

$$\boldsymbol{\mu}_{\theta_k|\theta_{k-1}} = E[\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}] = \boldsymbol{A}_k \boldsymbol{\mu}_{\theta_{k-1}}$$

and $$\boldsymbol{C}_{\theta_k|\theta_{k-1}} = E[(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k})((\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k})^T \mid \boldsymbol{\theta}_{k-1}] = \boldsymbol{A}_k \boldsymbol{C}_{\theta_{k-1}} \boldsymbol{A}_k^T + \boldsymbol{C}_{\varepsilon_k},$$

where $\boldsymbol{\mu}_{\theta_{k-1}}$ and $\boldsymbol{C}_{\theta_{k-1}}$ are the mean and covariance of $\boldsymbol{\theta}_{k-1}$. In other words, the dynamical equation allows us to predict the density of $\boldsymbol{\theta}_k$ given the previous one at $\boldsymbol{\theta}_{k-1}$.

(5 marks)

ii) In (6-2), the measurement $\boldsymbol{x}_k$ is taken so as to correct or update the density of $\boldsymbol{\theta}_k$ after the prediction from the dynamical equation. Let

$$\boldsymbol{z}_k = \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{y}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_k \boldsymbol{\theta}_k + \boldsymbol{w}_k \\ \boldsymbol{\theta}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_k & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_k \\ \boldsymbol{w}_k \end{bmatrix}.$$

As $\boldsymbol{z}_k$ is a linear transformation of a Gaussian vector $[\boldsymbol{\theta}_k^T, \boldsymbol{w}_k^T]^T$, it too is Gaussian distributed. Therefore, the mean and covariance (and hence the density) of $\boldsymbol{\theta}_k$ can again be updated using a classical result of Gaussian distribution relating its component vectors $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$ as follows

(6-3) $E[\boldsymbol{y}_k \mid \boldsymbol{x}_k] = E[\boldsymbol{y}_k] + \boldsymbol{C}_{y_k x_k} \boldsymbol{C}_{x_k x_k}^{-1} (\boldsymbol{x}_k - E[\boldsymbol{x}_k])$

(6-4) $\boldsymbol{C}_{y_k|x_k} = \boldsymbol{C}_{y_k y_k} - \boldsymbol{C}_{y_k x_k} \boldsymbol{C}_{x_k x_k}^{-1} \boldsymbol{C}_{x_k y_k}$,

where $\boldsymbol{C}_{ab} = E[(\boldsymbol{a} - E[\boldsymbol{a}])(\boldsymbol{b} - E[\boldsymbol{b}])^T]$, and $\boldsymbol{y}_k$ is now identified as $\boldsymbol{\theta}_k$. Due to the dynamical equation, the mean and covariance of $\boldsymbol{y}_k$, i.e. $\boldsymbol{\theta}_k$, are now given by $\boldsymbol{\mu}_{\theta_k|\theta_{k-1}}$ and $\boldsymbol{C}_{\theta_k|\theta_{k-1}}$ as derived in part (i) above.

Using (6-2), show that

1) $E[\boldsymbol{x}_k] = \boldsymbol{H}_k E[\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}] = \boldsymbol{H}_k \boldsymbol{\mu}_{\theta_k|\theta_{k-1}}$.

2) $\boldsymbol{C}_{\boldsymbol{x}_k} = E[(\boldsymbol{x}_k - E(\boldsymbol{x}_k))(\boldsymbol{x}_k - E(\boldsymbol{x}_k)^T \mid \boldsymbol{\theta}_{k-1}] = \boldsymbol{H}_k \boldsymbol{C}_{\theta_k|\theta_{k-1}} \boldsymbol{H}_k^T + \boldsymbol{C}_{w_k}$,

3) $\boldsymbol{C}_{\theta_k \boldsymbol{x}_k} = E[(\boldsymbol{\theta}_k - E(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}))(\boldsymbol{x}_k - E(\boldsymbol{x}_k))^T \mid \boldsymbol{\theta}_{k-1}] = \boldsymbol{C}_{\theta_k|\theta_{k-1}} \boldsymbol{H}_k^T$.

(10 marks)

iii) By substituting these results and those in part (i) into (6-3) and (6-4), show that

$$\boldsymbol{\mu}_{\theta_k} \equiv E(\boldsymbol{\theta}_k \mid \boldsymbol{x}_k) = \boldsymbol{\mu}_{\theta_k|\theta_{k-1}} + \boldsymbol{K}_k \boldsymbol{e}_k$$

and $\boldsymbol{C}_{\theta_k} \equiv \boldsymbol{C}_{\theta_k|x} = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)\boldsymbol{C}_{\theta_k|\theta_{k-1}}$,

where $\boldsymbol{\mu}_{\theta_k|\theta_{k-1}} = \boldsymbol{A}_k \boldsymbol{\mu}_{\theta_{k-1}}$, $\boldsymbol{K}_k = \boldsymbol{C}_{\theta_k|\theta_{k-1}} \boldsymbol{H}_k^T (\boldsymbol{H}_k \boldsymbol{C}_{\theta_k|\theta_{k-1}} \boldsymbol{H}_k^T + \boldsymbol{C}_{w_k})^{-1}$,

$\boldsymbol{C}_{\theta_k|\theta_{k-1}} = \boldsymbol{A}_k \boldsymbol{C}_{\theta_{k-1}} \boldsymbol{A}_k^T + \boldsymbol{C}_{\varepsilon_k}$ and $\boldsymbol{e}_k = (\boldsymbol{x}_k - \boldsymbol{H}_k \boldsymbol{\mu}_{\theta_k|\theta_{k-1}})$ is the error vector in predicting the measurement $\boldsymbol{x}_k$ from the dynamical equation. The process can be repeated over time and is known as the discrete Kalman filter.

(3 marks)

iv) Consider the recursive estimation of the phase angle ϕ of a sinusoid embedded in zero-mean white Gaussian noise $\varepsilon[n]$ with known variance σ_ε^2

$$x[k] = A\cos(2\pi f_0 k + \phi) + \varepsilon[n],\ k=0,1,2,\ldots$$

The frequency f_0 is assumed to be known and A>0.

By writing $\cos(2\pi f_0 k + \phi) = \cos(2\pi f_0 k)\cos(\phi) - \sin(2\pi f_0 k)\sin(\phi)$ and define the state vector $\boldsymbol{\theta}_k = [A\cos(\phi), A\sin(\phi)]^T$, illustrate how the Kalman filter above can be used to estimate the phase angle recursively.

[Hint: you can assume that $\boldsymbol{A}_k = \boldsymbol{I}$ and $\boldsymbol{C}_{w_k} = \sigma_w^2$ is known.] **(6 marks)**

***** END OF PAPER *****

SOLUTION

Q1.

(a)

Consider a causal linear time-invariant (LTI) system with the following transfer function

$$H(z) = \frac{2 - 2.4z^{-1} - 0.4z^{-2}}{(1+0.5z^{-1})(1-0.8z^{-1})} = \frac{2 - 2.4z^{-1} - 0.4z^{-2}}{(1-0.3z^{-1} - 0.4z^{-2})}.$$

(i) [**3 marks**] y(n)=2x(n)-2.4x(n-1)-0.4x(n-2)+0.3y(n-1)+0.4y(n-2).

(ii) [**2 marks**] The poles are $p_1 = -0.5$, $p_2 = 0.8$.

(iii) [**6 marks**] The impulse response is the inverse z-transform of $H(z)$. First, we express $H(z)$ as its partial fraction expansion,

$$H(z) = 1 + \frac{1 - 2.1z^{-1}}{(1+0.5z^{-1})(1-0.8z^{-1})} = 1 + \frac{2}{(1+0.5z^{-1})} - \frac{1}{(1-0.8z^{-1})}.$$

Taking z-transform and noting that the system is causal, one gets

$$h(n) = \delta(n) + 2(-0.5)^n u(n) - (0.8)^n u(n).$$

(iv) [**2 mark**s] Yes, because all the poles are inside the unit circle. If the system is non-causal, then the region of convergence of any one of the poles 0.5 and 0.8 will extend inwards, and the ROC cannot cover the unit circle. Therefore, the system will become unstable.

(v) **[2 marks]** Determine the z-transform of $x(n) = e^{j(n\omega_0)}u(n)$, where $u(n)$ is the unit step sequence.

$x(n) = e^{j(n\omega_0)}u(n)$. The z-transform is: $X(z) = \dfrac{1}{1 - e^{j\omega_0} z^{-1}}$, |z|>1.

(vi) [**6 marks**] $Y(z) = [1 + \dfrac{2}{(1+0.5z^{-1})} - \dfrac{1}{(1-0.8z^{-1})}]\dfrac{1}{1-e^{j\omega_0} z^{-1}}$, ROC $|z|>1$

$$= [1 + \frac{2}{(1+0.5e^{-j\omega_0})} - \frac{1}{(1-0.8e^{-j\omega_0})}]\frac{1}{1-e^{j\omega_0} z^{-1}}$$

$$+ \frac{1}{(1+(0.5)^{-1}e^{j\omega_0})}\frac{2}{(1+0.5z^{-1})} - \frac{1}{(1-(0.8)^{-1}e^{j\omega_0})}\frac{1}{(1-0.8z^{-1})}$$

$$y(n) = H(e^{j\omega_0})e^{jn\omega_0}u(n) + \left[\frac{2}{(1+(0.5)^{-1}e^{j\omega_0})}(-0.5)^n u(n) - \frac{1}{(1-(0.8)^{-1}e^{j\omega_0})}(0.8)^n u(n)\right].$$

The first term on the RHS is the steady state response whereas the term inside the bracket is

b)

i) [**5 marks**] The required impulse response is equal to the inverse DT-FT of

$$H_{hp}(e^{j\omega}) = \begin{cases} 0 & 0 \le |\omega| \le \omega_c \\ e^{-j\omega M/2} & \omega_c \le |\omega| \le \pi. \end{cases}$$

Hence,

$$h(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_{hp}(e^{j\omega}) e^{jn\omega} d\omega = \frac{1}{2\pi}\int_{\omega_c}^{\pi} e^{-j\omega[(M/2)-n]} d\omega + \frac{1}{2\pi}\int_{-\pi}^{-\omega_c} e^{-j\omega[(M/2)-n]} d\omega$$

$$= \frac{-1}{2\pi j[(M/2)-n]}\{[e^{-j[(M/2)-n]\pi} - e^{-j[(M/2)-n]\omega_c}] + [e^{j[(M/2)-n]\omega_c} - e^{j[(M/2)-n]\pi}]\}$$

$$= \frac{\sin \pi[(M/2)-n]}{\pi[(M/2)-n]} - \frac{\sin \omega_c[(M/2)-n]}{\pi[(M/2)-n]}.$$

ii) [5 marks]

Given $\quad x(n) = u(-n-1) + \left(\frac{1}{2}\right)^n u(n),$

and $\quad X(z) = \frac{-1}{1-z^{-1}} + \frac{1}{1-\frac{1}{2}z^{-1}} \qquad \frac{1}{2} < |z| < 1$

To find $H(z)$, we simply use $Y(z) = \frac{-\frac{1}{2}z^{-1}}{(1-\frac{1}{2}z^{-1})(1+z^{-1})}$ to obtain

$$\begin{aligned} H(z) &= \frac{Y(z)}{X(z)} \\ &= \frac{-\frac{1}{2}z^{-1}}{(1-\frac{1}{2}z^{-1})(1+z^{-1})} \cdot \frac{(1-z^{-1})(1-\frac{1}{2}z^{-1})}{-\frac{1}{2}z^{-1}} \\ &= \frac{(1-z^{-1})}{(1+z^{-1})}. \end{aligned}$$

Since $H(z)$ is causal, ROC is $|z| > 1$.

Since one of the poles X(z), which limited the ROC of X(z) to be less than 1, is cancelled by the zero of $H(z)$, the ROC of $Y(z)$ is the region in the z-plane that satisfies the remaining two constraints $|z| > \frac{1}{2}$ and $|z| > 1$.

Therefore $\quad y(n) = -\frac{1}{3}\left(\frac{1}{2}\right)^n u(n) + \frac{1}{3}(-1)^n u(n)$.

Q2.

(i) [**4 marks**] The discrete-time specifications are

$$0.99 \le \left|H(e^{j\omega})\right| \le 1.01, \qquad 0 \le \omega \le 0.4\pi\,.$$

$$\left|H(e^{j\omega})\right| \le 0.001, \qquad 0.50\pi \le \omega \le \pi\,.$$

(ii) [**4 marks**] Use the minimum specifications, we have

$$\delta = 0.001$$

$$\Delta\omega = 0.1\pi$$

$$A = -20 \cdot \log_{10} \delta = 60 \text{ dB}$$

$$\beta = 0.1102(A - 8.7) = 5.65326$$

$M = \dfrac{A-8}{2.285\Delta\omega} = 72.438$. Hence choose M to be 73.

iii) [**5 marks**] The required filter length from the McClellan-Parks algorithm is

$N = \dfrac{-20\log_{10}\sqrt{0.01 \cdot 0.001} - 13}{14.6 \cdot 0.05} = 50.685$ Hence, choose N to be 51.

The McClellan-Parks design algorithm is more efficient and leads to a lower filter length. The design of the Kaiser window is very simple.

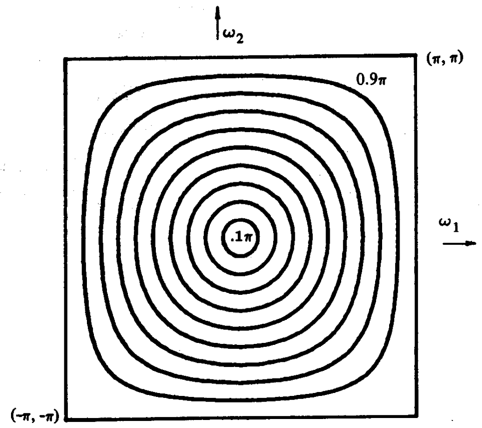
(b)

i) **[3 marks]** Show that $H(\pm\omega_2, \pm\omega_1) = H(\omega_1, \omega_2)$.

Since every term in $F(\omega_1, \omega_2) = \frac{1}{2}(-1 + \cos\omega_1 + \cos\omega_2 + \cos\omega_1 \cos\omega_2)$ is an even function, we have $H(\pm\omega_1, \pm\omega_2) = H(\omega_1, \omega_2)$. Further, $F(\omega_1, \omega_2)$ is symmetric in (ω_1, ω_2), hence $F(\omega_1, \omega_2) = F(\omega_2, \omega_1)$. Combining the two results gives

$$H(\pm\omega_2, \pm\omega_1) = H(\omega_1, \omega_2).$$

ii) [**4 marks**]



The passband should lie in the region enclosed by the contour labeled 0.4π. The stopband should lie in the region outside the contour labeled 0.5π.

The circular symmetry will degrade when the cutoff frequencies are increased towards π.

iii) [**2 marks**] Least squares method is very flexible and it can design filter with different desired response. Its disadvantage is that there are large sidelobes around transition band (discontinuities to be approximated). The McClellan transformation method only work for linear-phase filters with certain spectral support. Its advantage is its good performance and efficient implementation because the 1D prototype can be designed optimally by the Parks-McClellan method. The ripple is usually equiripple.

(c)

i) **[2 marks]**

$$H(e^{j\omega_1}, e^{j\omega_2}) = \sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} h(n_1, n_2) e^{-jn_1\omega_1} e^{-jn_2\omega_2}$$

$$= \sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} h(n_1, n_2)\cos(n_1\omega_1 + n_2\omega_2) - j\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} h(n_1, n_2)\sin(n_1\omega_1 + n_2\omega_2)$$

$$= \boldsymbol{h}^T[\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega})],$$

where $\boldsymbol{\omega} = [\omega_1, \omega_2]^T$, $\boldsymbol{h} = [h(0,0), h(0,1), ..., h(N_1 - 1, N_2 - 2), h(N_1 - 1, N_2 - 1)]^T$, $\boldsymbol{c}(\boldsymbol{\omega}) = [1, \cos(\omega_2), ..., \cos((N_1 - 1)\omega_1 + (N_2 - 1)\omega_2)]^T$ and $\boldsymbol{s}(\boldsymbol{\omega}) = [0, ..., \sin((N_1 - 1)\omega_1 + (N_2 - 1)\omega_2)]^T$.

ii) **[3 marks]** Show that the WLS function is given by

$$\mathrm{WLS}(\boldsymbol{h}) = \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega}) \,|\, H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}) \,|^2 d\boldsymbol{\omega}$$

$$= \boldsymbol{h}^T \boldsymbol{Q} \boldsymbol{h} - 2\boldsymbol{h}^T \boldsymbol{g} + c \,,$$

where $\boldsymbol{Q} = \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega})(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega}))(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega}))^H \, d\boldsymbol{\omega}$,

$$\boldsymbol{g} = \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega}) \operatorname{Re}\{(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega})) H_d^*(e^{j\omega_1}, e^{j\omega_2})\} d\boldsymbol{\omega} \,,$$

$c = \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega}) \,|\, H_d(e^{j\omega_1}, e^{j\omega_2}) \,|^2 \, d\boldsymbol{\omega}$ and superscript ^{H} stands for conjugate transpose.

[Hint:

$$\begin{aligned} &|\, H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}) \,|^2 \\ &\quad = (H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}))(H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}))^* \,] \end{aligned}$$

$$\begin{aligned}
&\mathrm{WLS}(\boldsymbol{h}) \\
&= \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega})(H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}))(H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}))^* \, d\boldsymbol{\omega} \\
&= \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega})(|\, H(e^{j\omega_1}, e^{j\omega_2}) \,|^2 - 2\operatorname{Re}\{H(e^{j\omega_1}, e^{j\omega_2}) H_d^*(e^{j\omega_1}, e^{j\omega_2})\} + |\, H_d(e^{j\omega_1}, e^{j\omega_2}) \,|^2) d\boldsymbol{\omega} \\
&= \boldsymbol{h}^T [\int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega})(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega}))(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega}))^H \, d\boldsymbol{\omega}] \boldsymbol{h} \\
&\qquad - 2\boldsymbol{h}^T [\int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega}) \operatorname{Re}\{(\boldsymbol{c}(\boldsymbol{\omega}) - j\boldsymbol{s}(\boldsymbol{\omega})) H_d^*(e^{j\omega_1}, e^{j\omega_2})\} d\boldsymbol{\omega}] \\
&\qquad + \int_{\Omega_p \cup \Omega_s} W(\boldsymbol{\omega}) \,|\, H_d(e^{j\omega_1}, e^{j\omega_2}) \,|^2 \, d\boldsymbol{\omega} \\
&= \boldsymbol{h}^T \boldsymbol{Q} \boldsymbol{h} - 2\boldsymbol{h}^T \boldsymbol{g} + c \,.
\end{aligned}$$

iii) **(6 marks)**

In norm constrained digital filter design, the 2-norm of the filter coefficients, $\|\boldsymbol{h}\|_2^2 = \boldsymbol{h}^T \boldsymbol{h}$, are also minimized. The objective function to be minimized is thus

$$E(\boldsymbol{h}) = \mathrm{WLS}(\boldsymbol{h}) + \lambda \|\boldsymbol{h}\|_2^2$$

where λ is a constant to control the importance of the two terms above.

By discretizing ω_1, ω_2 uniformly in $\omega_1, \omega_2 \in \Omega_E$ into M points, $\boldsymbol{\omega}_i$ for i=1,…,M, show that the above problem can be written as:

$$\min_{\boldsymbol{h},\delta} \quad \delta$$

$$\text{subject to} \quad \delta_p - \{\alpha_R^2(\boldsymbol{\omega}_i) + \alpha_I^2(\boldsymbol{\omega}_i)\}^{1/2} \geq 0, \quad i = 1, \ldots, M,$$

$$\| \tilde{\boldsymbol{L}}^T \boldsymbol{h} - \tilde{\boldsymbol{L}}^{-1} \boldsymbol{g} \|_2 \leq \delta$$

$$\alpha_R(\boldsymbol{\omega}) = W(\boldsymbol{\omega}) \cdot \{\boldsymbol{h}^T \boldsymbol{c}(\boldsymbol{\omega}) - \mathrm{Re}[H_d(e^{j\omega_1}, e^{j\omega_2})]\},$$

$$\alpha_I(\boldsymbol{\omega}) = W(\boldsymbol{\omega}) \cdot \{\boldsymbol{h}^T \boldsymbol{s}(\boldsymbol{\omega}) + \mathrm{Im}[H_d(e^{j\omega_1}, e^{j\omega_2})]\}.$$

Express $\tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^T$ in terms of $\boldsymbol{Q}$.

[Hint: If $\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^T$ where $\tilde{\boldsymbol{L}}$ is some nonsingular upper triangular matrix, then

$$\min_{\boldsymbol{h}} \quad \boldsymbol{h}^T \tilde{\boldsymbol{Q}} \boldsymbol{h} - 2\boldsymbol{h}^T \boldsymbol{g} + c$$

is equivalent to

$$\min_{\boldsymbol{h}} \| \tilde{\boldsymbol{L}}^T \boldsymbol{h} - \tilde{\boldsymbol{L}}^{-1} \boldsymbol{g} \|_2^2,$$

where the $L2$ norm of a vector $\boldsymbol{x}$ is given by $\| \boldsymbol{x} \|_2^2 = \boldsymbol{x}^T \boldsymbol{x}$. Find the new $\tilde{\boldsymbol{Q}}$ of $E(\boldsymbol{h})$ using Part (ii).]

Since

$$\| \boldsymbol{L}^T \boldsymbol{h} - \boldsymbol{L}^{-1} \boldsymbol{g} \|_2^2 = \boldsymbol{h}^T \boldsymbol{L}\boldsymbol{L}^T \boldsymbol{h} - 2\boldsymbol{h}^T \boldsymbol{L}\boldsymbol{L}^{-1} \boldsymbol{g} + \boldsymbol{g}^T \boldsymbol{L}^{-T} \boldsymbol{L}^{-1} \boldsymbol{g} = \boldsymbol{h}^T \boldsymbol{Q} \boldsymbol{h} - 2\boldsymbol{h}^T \boldsymbol{g} + \| \boldsymbol{L}^{-1} \boldsymbol{g} \|_2^2$$

and both $\| \boldsymbol{L}^{-1} \boldsymbol{g} \|_2^2$ and c are constants independent of $\boldsymbol{h}$, hence

$\min_{\boldsymbol{h}} \ \mathrm{WLS}(\boldsymbol{h}) = \boldsymbol{h}^T \boldsymbol{Q} \boldsymbol{h} - 2\boldsymbol{h}^T \boldsymbol{g} + c$ is equivalent to $\min_{\boldsymbol{h}} \| \boldsymbol{L}^T \boldsymbol{h} - \boldsymbol{L}^{-1} \boldsymbol{g} \|_2^2$.

Similarly, $\min_{\boldsymbol{h}} \ \mathrm{WLS}(\boldsymbol{h}) + \lambda \| \boldsymbol{h} \|_2^2 = \boldsymbol{h}^T (\boldsymbol{Q} + \lambda \boldsymbol{I}) \boldsymbol{h} - 2\boldsymbol{h}^T \boldsymbol{g} + c$ is equivalent to $\min_{\boldsymbol{h}} \| \tilde{\boldsymbol{L}}^T \boldsymbol{h} - \tilde{\boldsymbol{L}}^{-1} \boldsymbol{g} \|_2^2$, where $\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^T = (\boldsymbol{Q} + \lambda \boldsymbol{I})$.

Let δ^2 be an upper bound of $E(\boldsymbol{h})$. The WLS minimization can be achieved by minimizing δ such that $\| \tilde{\boldsymbol{L}}^T \boldsymbol{h} - \tilde{\boldsymbol{L}}^{-1} \boldsymbol{g} \|_2 \leq \delta$. Together with the other peak magnitude constraints, one gets:

$$\min_{\boldsymbol{h},\delta} \delta$$

$$\text{subject to } W(\boldsymbol{\omega}) | H(e^{j\omega_1}, e^{j\omega_2}) - H_d(e^{j\omega_1}, e^{j\omega_2}) | \leq \delta_p, \text{ for } \omega \in \Omega_E$$

$$\| \tilde{\boldsymbol{L}}^T \boldsymbol{h} - \tilde{\boldsymbol{L}}^{-1} \boldsymbol{g} \|_2 \leq \delta.$$

By discretizing $\boldsymbol{\omega}$ uniformly into M points in the passband and stopband edges $\omega \in \Omega_E$, $\boldsymbol{\omega}_i$ for i=1,…,M, the above problem can be written as

$$\min_{\boldsymbol{h},\delta} \delta$$

subject to $W(\boldsymbol{\omega}_i) | H(e^{j\omega_i}) - H_d(e^{j\omega_i}) | \le \delta_p$, i=1,…,M.

$$\| \tilde{\boldsymbol{L}}^T \boldsymbol{h} - \tilde{\boldsymbol{L}}^{-1} \boldsymbol{g} \|_2 \le \delta$$

Now $W(\boldsymbol{\omega}_i)[H(e^{j\omega_i}) - H_d(e^{j\omega_i})] = W(\boldsymbol{\omega}_i)[\boldsymbol{h}^T[\boldsymbol{c}(\boldsymbol{\omega}_i) - j\boldsymbol{s}(\boldsymbol{\omega}_i)] - H_d(e^{j\omega_i})]$ $= \alpha_R(\boldsymbol{\omega}_i) + j\alpha_I(\boldsymbol{\omega}_i)$,

where $\alpha_R(\boldsymbol{\omega}) = W(\boldsymbol{\omega}) \cdot \{\boldsymbol{h}^T \boldsymbol{c}(\boldsymbol{\omega}) - \mathrm{Re}[H_d(e^{j\omega_i})]\}$,

and $\alpha_I(\boldsymbol{\omega}) = W(\boldsymbol{\omega}) \cdot \{\boldsymbol{h}^T \boldsymbol{s}(\boldsymbol{\omega}) + \mathrm{Im}[H_d(e^{j\omega_i})]\}$. Therefore, the constraints become

$W(\boldsymbol{\omega}_i) | H(e^{j\omega_i}) - H_d(e^{j\omega_i}) | \le \delta_p \Leftrightarrow \delta_p - \{\alpha_R^2(\boldsymbol{\omega}_i) + \alpha_I^2(\boldsymbol{\omega}_i)\}^{1/2} \ge 0$. Hence, the problem becomes

$$\min_{\boldsymbol{h},\delta} \quad \delta$$

$$\text{subject to} \quad \delta_p - \{\alpha_R^2(\boldsymbol{\omega}_i) + \alpha_I^2(\boldsymbol{\omega}_i)\}^{1/2} \ge 0, \quad i = 1, \ldots, M,$$

$$\| \tilde{\boldsymbol{L}}^T \boldsymbol{h} - \tilde{\boldsymbol{L}}^{-1} \boldsymbol{g} \|_2 \le \delta,$$

$$\alpha_R(\boldsymbol{\omega}) = W(\boldsymbol{\omega}) \cdot \{\boldsymbol{h}^T \boldsymbol{c}(\boldsymbol{\omega}) - \mathrm{Re}[H_d(e^{j\omega})]\}, \ \alpha_I(\boldsymbol{\omega}) = W(\boldsymbol{\omega}) \cdot \{\boldsymbol{h}^T \boldsymbol{s}(\boldsymbol{\omega}) + \mathrm{Im}[H_d(e^{j\omega})]\}.$$

Q.3

a)

i) [**4 marks**]

The output of the system is given by

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k).$$

Taking expectation on both sides, one gets

$$E[y(n)] = E\left[\sum_{k=-\infty}^{\infty} h(k)x(n-k)\right]$$

$$= \sum_{k=-\infty}^{\infty} h(k)E[x(n-k)] = \mu_x \cdot \sum_{k=-\infty}^{\infty} h(k) = \mu_x \cdot H(1),$$

because $x(n)$ be a wide-sense stationary random process with mean μ_x.

$$R_{yy}(k) = E[y(n)y(n+k)] = E\left[\sum_{i=-\infty}^{\infty} h(i)x(n-i) \cdot \sum_{l=-\infty}^{\infty} h(l)x(n+k-l)\right]$$

$$= \sum_{i=-\infty}^{\infty} h(i) \sum_{l=-\infty}^{\infty} h(l) E[x(n-i)x(n+k-l)]$$

$$= \sum_{i=-\infty}^{\infty} h(i) \sum_{l=-\infty}^{\infty} h(l) R_{xx}(k-l+i)$$

$$= \sum_{i=-\infty}^{\infty} h(-i)(R_{xx} * h)(k-i) = R_{xx}(k) * h(k) * h(-k).$$

Hence, $y(n)$ is also a wide-sense stationary process with mean μ_y and autocorrelation $R_{yy}(k)$ given by,

$$\mu_y = H(1) \cdot \mu_x,$$

$$\text{and} \quad R_{yy}(k) = R_{xx}(k) * h(k) * h(-k).$$

ii) [**3 marks**]

Taking the DTFT of $R_{yy}(k)$, we get the the Power Spectral Density (PSD) of $y(n)$ as

$$S_{yy}(e^{j\omega}) = DTFT[R_{yy}(k)] = DTFT[R_{xx}(k) * h(k) * h(-k)].$$

Using the convolution theorem, we have

$$S_{yy}(e^{j\omega}) = DTFT[R_{xx}(k) * h(k) * h(-k)]$$

$$= S_{xx}(e^{j\omega}) H(e^{j\omega}) H^*(e^{j\omega}) = S_{xx}(e^{j\omega}) \left| H(e^{j\omega}) \right|^2.$$

b)

i) [**3 marks]**

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

ii) [**5 marks**] The transfer function between $\xi_{a_i}(n)$ and $e_{a_i}(n)$ is $H(z)$.

Let its impulse response be $h(n)$. Hence

$$\xi_{a_i}(n) = e_{a_i}(n) * h(n) = \sum_{k=0}^{\infty} e_{a_i}(n-k) h(k). \qquad \text{(3-1a)}$$

Consider

$$E[\{\xi_{a_1}(n)+\xi_{a_1}(n)+\sum_{j=0}^{2}e_{b_j}(n)\}^2]=E[\xi_{a_1}^2(n)]+E[\xi_{a_2}^2(n)]+\sum_{j=0}^{2}E[e_{b_j}^2(n)]$$

$$+2E[\xi_{a_1}(n)\xi_{a_2}(n)]+2E[\sum_{j=0}^{2}\xi_{a_1}(n)e_{b_j}(n)]+2E[\sum_{j=0}^{2}\xi_{a_2}(n)e_{b_j}(n)]+2E[\sum_{0\le j\ne k\le 2}e_{b_j}(n)e_{b_k}(n)],$$

Since $e_{a_i}(n)$ and $e_{b_i}(n)$ are independent and are of zero means, the terms $\xi_{a_i}(n)$ and $e_{b_i}(n)$ are independent and are of zero means. The terms $E[\sum_{j=0}^{2}\xi_{a_1}(n)e_{b_j}(n)]$, $E[\sum_{j=0}^{2}\xi_{a_2}(n)e_{b_j}(n)]$, and $E[\sum_{0\le j\ne k\le 2}e_{b_j}(n)e_{b_k}(n)]$ are equal to zero. To show that $E[\xi_{a_1}(n)\xi_{a_2}(n)]=0$, we have after using (3-1) the following

$$E[\xi_{a_1}(n)\xi_{a_2}(n)] = E\left[\sum_{k=0}^{\infty}e_{a_1}(n-k)h(k)\cdot\sum_{m=0}^{\infty}e_{a_2}(n-m)h(n)\right]$$

$$=\left[\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}E[e_{a_1}(n-k)e_{a_2}(n-m)]h(k)h(n)\right]=0$$

(because $e_{a_1}(n)$ and $e_{a_2}(n)$ are uncorrelated).

iii) [**3 marks**] The variances of $e_{a_i}(n)$ is given by

$$\sigma_{e_{a_i}}^2 = \int_{-2^{-(B+1)}}^{2^{-(B+1)}} x^2\cdot p_{a_i}(x)dx,$$

where $p_{a_i}(x)$ is the probability density function of $e_{a_i}(n)$ in the interval $[-\frac{1}{2}2^{-B_a},\frac{1}{2}2^{-B_a}]$. Since the quantization error is uniformly distributed in the interval $[-\frac{1}{2}2^{-B_a},\frac{1}{2}2^{-B_a}]$, we have $p_{a_i}(x)=\frac{1}{2^{-B_a}}$. Hence

$$\sigma_{e_{a_i}}^2 = \frac{1}{2^{-B_a}}\int_{-2^{-(B_a+1)}}^{2^{-(B_a+1)}} x^2 dx = \frac{1}{3\cdot 2^{-B_a}}x^3\Big|_{-2^{-(B_a+1)}}^{2^{-(B_a+1)}}$$

$$=\frac{2\cdot 2^{-3(B_a+1)}}{3\cdot 2^{-B_a}}=\frac{2^{-2B_a}}{12}.$$

Similarly, we have $\sigma^2_{e_{b_i}} = \frac{2^{-2B_b}}{12}$.

Since $\sigma^2_{e_{a_i}} = \frac{2^{-2B_a}}{12}; \sigma^2_{e_{b_j}} = \frac{2^{-2B_b}}{12}$ and the quantization errors are white, $P_{a_i}(\omega) = P_a; P_{b_j}(\omega) = P_b$ are constants. Hence,

$$\sigma^2_{e_{a_i}} = = \frac{2^{-2B_a}}{12} \frac{1}{2\pi} \int_{-\pi}^{\pi} P_a d\omega = P_a$$

$$\sigma^2_{e_{b_j}} = \frac{2^{-2B_b}}{12} = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_b d\omega = P_b.$$

Thus, we have $P_{a_i}(\omega) = \frac{2^{-2B_a}}{12}$. $P_{b_j}(\omega) = \frac{2^{-2B_b}}{12}$

iv) [**5 marks**] The power spectral density of $\xi_{a_i}(n)$ is given by

$$P_{\xi_a}(\omega) = P_{a_i}(\omega) \left| H(e^{j\omega}) \right|^2$$

$$= \frac{2^{-2B_a}}{12} \left| H_e(e^{j\omega}) \right|^2.$$

Similarly, we have

$$P_{\xi_{b_j}}(\omega) = \frac{2^{-2B_b}}{12}.$$

The variances of $\xi_{a_i}(n)$ are thus

$$\sigma^2_{\xi_{a_i}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{2^{-2B_a}}{12} \left| H(e^{j\omega}) \right|^2 d\omega.$$

Similar,

$$\sigma^2_{\xi_{b_j}} = \frac{2^{-2B_b}}{12}.$$

Using the Parseval's theorem

$$\sum_{n=-\infty}^{\infty} |h(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| H(e^{j\omega}) \right|^2 d\omega,$$

we have

$$\sigma^2_{\xi_{a_i}} = \frac{2^{-2B_a}}{12} \sum_{n=-\infty}^{\infty} |h(n)|^2 .$$

Finally, we have

$$\sigma^2_{\xi} = 3\frac{2^{-2B_b}}{12} + 2\frac{2^{-2B_a}}{12} \sum_{n=-\infty}^{\infty} | h[n] |^2 .$$

c) Consider a zero mean real random vector $\boldsymbol{x} = [x(1),..., x(M)]^T$ with covariance matrix $\boldsymbol{R}_{xx} = E[\boldsymbol{x}\boldsymbol{x}^T]$. The total variance of $\boldsymbol{x}$ is $\sigma_x^2 = E[\boldsymbol{x}^T\boldsymbol{x}] = \sum_{i=1}^{M} E[x^2(i)] = \sum_{i=1}^{M} \sigma^2_{x(i)}$. Consider the transformed vector $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} = [y(1),..., y(M)]^T$, where $\boldsymbol{C}$ is an orthogonal transformation.

i) **[2 marks]** Show that the total variance of the transformation vector $\boldsymbol{y}$ $\sigma_y^2 = E[\boldsymbol{y}^T\boldsymbol{y}] = \sum_{i=1}^{M} \sigma^2_{y(k)}$ is also σ_x^2.

$$\sigma_y^2 = E[\boldsymbol{y}^T\boldsymbol{y}] = E[\boldsymbol{x}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{x}] = E[\boldsymbol{x}^T\boldsymbol{x}] = \sigma_x^2,$$

since $\boldsymbol{C}$, $\boldsymbol{C}^T\boldsymbol{C} = \boldsymbol{I}$.

ii) **[3 marks]** Suppose that the k-th transformed coefficient, $y(k)$, is to be quantized by a uniform quantizer with b_k number of bits. The total number of bits, b, for representing each realization of the random vector $\boldsymbol{y}$ is then given by $b = \sum_{k=1}^{M} b_k$. Let the quantization error be $q(k) = y(k) - \hat{y}(k)$ where $\hat{y}(k)$ is the corresponding quantized value of the transformed coefficient. Similarly, let the corresponding error in the original domain be $\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{C}^T\hat{\boldsymbol{y}}$ where $\boldsymbol{e} = [e(1),..., e(M)]^T$ and $\hat{\boldsymbol{y}} = [\hat{y}(1),..., \hat{y}(M)]^T$.

Show that $\sigma_e^2 = \sum_{k=1}^{M} \sigma^2_{q(k)} = \sum_{k=1}^{M} E[q^2(k)]$.

Since $\boldsymbol{e} = \boldsymbol{C}^T(\boldsymbol{y} - \hat{\boldsymbol{y}})$ and $\boldsymbol{q} = \boldsymbol{y} - \hat{\boldsymbol{y}}$, one has

$$\boldsymbol{e} = \boldsymbol{C}^T\boldsymbol{q} .$$

Using the result in part i), $\sigma_e^2 = \sigma_q^2 = \sum_{k=1}^{M-1} \sigma^2_{q(k)}$.

iii) **[3 marks]** Assume that all the quantizers can be modeled by the following formula:

$$\sigma_{q(k)}^2 = c \cdot 2^{-2b_k} \sigma_{y(k)}^2 \qquad (*)$$

where c is a constant.

Show that $\sigma_e^2 \geq cM \cdot 2^{-\frac{2}{M}b} \cdot \left(\prod_{k=1}^{M} \sigma_{y(k)}^2\right)^{1/M}$.

[Hint: Let $a_k = 2^{-2b_k} \sigma_{y(k)}^2 \geq 0$) and substitute (*) into the result of part (ii). Then, use the fact that arithmetic mean is greater or equal to the geometric mean for nonnegative number a_k, i.e.

$$\frac{1}{M}\sum_{k=1}^{M} a_k \geq \left(\prod_{k=1}^{M} a_k\right)^{1/M},$$

with equality if and only if $a_1 = ... = a_M$.]

Using the result in part (ii), the total distortion is

$$\sigma_e^2 = c\sum_{k=1}^{M} 2^{-2b_k} \sigma_{y(k)}^2 .$$

Let $a_k = 2^{-2b_k} \sigma_{y(k)}^2 \geq 0$, one gets using the fact that arithmetic mean is greater or equal to the geometric mean for nonnegative number, the following:

$$\sigma_e^2 = c\sum_{k=1}^{M} 2^{-2b_k} \sigma_{y(k)}^2 \geq cM\left(\prod_{k=1}^{M} 2^{-2b_k} \sigma_{y(k)}^2\right)^{1/M} = cM\left(\prod_{k=1}^{M} 2^{-2b_k}\right)^{1/M} \left(\prod_{k=1}^{M} \sigma_{y(k)}^2\right)^{1/M}$$

$$= cM \cdot 2^{-\frac{1}{M}\sum_{k=1}^{M} 2b_k} \cdot \left(\prod_{k=1}^{M} \sigma_{y(k)}^2\right)^{1/M} = cM \cdot 2^{-\frac{2}{M}b} \cdot \left(\prod_{k=1}^{M} \sigma_{y(k)}^2\right)^{1/M},$$

where we have used $b = \sum_{k=1}^{M} b_k$.

iv) **[2 marks]** Since the right hand side of the inequality in part (iii) depends only on b and $\left(\prod_{k=1}^{M} \sigma_{y(k)}^2\right)^{1/M}$, but not on b_k, it can be made equal by choosing b_k such that $a_1 = ... = a_M$. This requires $a_k = \sigma_{q(k)}^2$, k=1,…,M, to be identical. Let denote the corresponding minimum average distortion in time domain by $\sigma_{e,T\min}^2$. If PCM is employed to quantize the components of vector $\boldsymbol{x}$ directly, the number of bits per coefficient used will be b/M. The corresponding average distortion of $x(k)$ is $\sigma_{x(k)}^2 = c \cdot 2^{-2b/M} \sigma_{x(k)}^2$. Denote the total average distortion by $\sigma_{e,PCM}^2 = \sum_{k=1}^{M} \sigma_{x(k)}^2$. Using the results in parts (i) to (iii), show that the coding gain of using transformation over PCM is given by

$$\frac{\sigma^2_{e,PCM}}{\sigma^2_{e,T\min}} = \frac{\frac{1}{M}\sum_{k=1}^{M}\sigma^2_{y(k)}}{\left(\prod_{k=1}^{M}\sigma^2_{y(k)}\right)^{1/M}} .$$

The coding gain over PCM is

$$\frac{\sigma^2_{e,PCM}}{\sigma^2_{e,T\min}} = \frac{c2^{-(2b/M)}\cdot\sum_{k=1}^{M}\sigma^2_{x(k)}}{cM\cdot 2^{-(2b/M)}\cdot\left(\prod_{k=1}^{M}\sigma^2_{y(k)}\right)^{1/M}} = \frac{\frac{1}{M}\sum_{k=1}^{M}\sigma^2_{y(k)}}{\left(\prod_{k=1}^{M}\sigma^2_{y(k)}\right)^{1/M}} ,$$

where we have used the fact that $\sum_{k=1}^{M}\sigma^2_{x(k)} = \sum_{k=1}^{M}\sigma^2_{y(k)}$.

Q.4

i) **[4 marks]** The MSE is

$$\begin{aligned}
\xi &= E[e^2(n)] = E[(d(n)-y(n))^2] \\
&= E[(d(n)-W^T\boldsymbol{X}_n)^2] \\
&= E[d^2(n)] + E[\boldsymbol{W}^T\boldsymbol{X}_n\boldsymbol{X}_n^T\boldsymbol{W}] - 2E[d(n)\boldsymbol{X}_n^T\boldsymbol{W}] \\
&= E[d^2(n)] + \boldsymbol{W}^T E[\boldsymbol{X}_n\boldsymbol{X}_n^T]\boldsymbol{W} - 2E[d(n)\boldsymbol{X}_n^T]\boldsymbol{W} \\
&= r_{dd}(0) + \boldsymbol{W}^T\boldsymbol{R}\boldsymbol{W} - 2\boldsymbol{P}^T\boldsymbol{W} ,
\end{aligned}$$

ii) **[3 marks]** Differentiating ξ with respect to the weight vector, we obtain

$$\nabla = \frac{\partial\xi}{\partial\boldsymbol{W}} = 2\boldsymbol{R}\boldsymbol{W} - 2\boldsymbol{P} .$$

The optimal weight vector $\boldsymbol{W}*$ is obtained by setting the gradient to zero. This yields the following normal equation

$$\boldsymbol{R}\boldsymbol{W}* = \boldsymbol{P} ,$$

and $\boldsymbol{W}* = \boldsymbol{R}^{-1}\boldsymbol{P}$.

iii) **[6 marks]** iii) In recursive least squares (RLS) adaptive filtering algorithm, the correlation matrix $\boldsymbol{R}(n)$ and correlation vector $\boldsymbol{P}(n)$ at the *n-th* time instant are respectively estimated recursively as:

$$\boldsymbol{R}(n) = \lambda\boldsymbol{R}(n-1) + \boldsymbol{X}_n\boldsymbol{X}_n^T \text{ and } \boldsymbol{P}(n) = \lambda\boldsymbol{P}(n-1) + d(n)\boldsymbol{X}_n ,$$

where $0 < \lambda < 1$ is a forgetting factor, $\boldsymbol{R}(0) = \delta\boldsymbol{I}$ and $\boldsymbol{P}(0) = \boldsymbol{0}$ with δ a small positive number.

[1 marks] Show that $\boldsymbol{R}(n) = \lambda^n \delta \boldsymbol{I} + \tilde{\boldsymbol{X}}_n^T \tilde{\boldsymbol{X}}_n$ and $\boldsymbol{P}(n) = \tilde{\boldsymbol{X}}_n^T \boldsymbol{d}_n$,

where $\tilde{\boldsymbol{X}}_n^T = [\lambda^{(n-1)/2} \boldsymbol{X}_1, \lambda^{(n-2)/2} \boldsymbol{X}_2, ..., \lambda^{1/2} \boldsymbol{X}_{n-1}, \boldsymbol{X}_n]$ and $\boldsymbol{d}_n = [d(1), ..., d(n)]^T$.

Obviously, by multiplying out $\tilde{\boldsymbol{X}}_n^T \tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{X}}_n^T \boldsymbol{d}_n$, one gets the terms $\sum_{i=1}^{n} \lambda^{n-i} \boldsymbol{X}_i \boldsymbol{X}_i^T$ and $\sum_{i=1}^{n} \lambda^{n-i} \boldsymbol{X}_i d(i)$, respectively. Since $\boldsymbol{R}(0) = \delta \boldsymbol{I}$, it is equal to $\lambda^n \delta \boldsymbol{I}$ after n time steps. This yields $\boldsymbol{R}(n) = \lambda^n \delta \boldsymbol{I} + \tilde{\boldsymbol{X}}_n^T \tilde{\boldsymbol{X}}_n$.

[2 marks] Show that $\boldsymbol{R}(n) = \boldsymbol{C}_n^T \boldsymbol{C}_n$ and $\boldsymbol{P}(n) = \boldsymbol{C}_n^T \tilde{\boldsymbol{d}}_n$, where $\boldsymbol{C}_n = \begin{bmatrix} \sqrt{\delta \lambda^n} \boldsymbol{I} \\ \tilde{\boldsymbol{X}}_n \end{bmatrix}$ and $\tilde{\boldsymbol{d}}_n = [\boldsymbol{0}_L^T, \boldsymbol{d}_n^T]^T$, where $\boldsymbol{0}_L$ is a zero column vector of size L.

$$\boldsymbol{C}_n^T \boldsymbol{C}_n = \begin{bmatrix} \sqrt{\delta \lambda^n} \boldsymbol{I} & \tilde{\boldsymbol{X}}_n^T \end{bmatrix} \begin{bmatrix} \sqrt{\delta \lambda^n} \boldsymbol{I} \\ \tilde{\boldsymbol{X}}_n \end{bmatrix} = \lambda^n \delta \boldsymbol{I} + \tilde{\boldsymbol{X}}_n^T \tilde{\boldsymbol{X}}_n = \boldsymbol{R}(n).$$

$$\boldsymbol{C}_n^T \tilde{\boldsymbol{d}}_n = \begin{bmatrix} \sqrt{\delta \lambda^n} \boldsymbol{I} & \tilde{\boldsymbol{X}}_n^T \end{bmatrix} \begin{bmatrix} 0 \\ \boldsymbol{d}_n \end{bmatrix} = \tilde{\boldsymbol{X}}_n^T \boldsymbol{d}_n = \boldsymbol{P}(n)$$

[3 marks] Suppose $\boldsymbol{C}_n$ can be factored recursively using the QR decomposition so that $\boldsymbol{Q}_n^T [\boldsymbol{C}_n, \tilde{\boldsymbol{d}}_n] = [\boldsymbol{R}_n, \boldsymbol{d}_n^{'}]$, where, $\boldsymbol{Q}_n$ is some orthogonal matrix and $\boldsymbol{R}_n$ is an (LxL) upper triangular matrix. Show that the normal equation

$$\boldsymbol{R}(n) \boldsymbol{W}(n) = \boldsymbol{P}(n)$$

can be simplified to that $\boldsymbol{R}_n \boldsymbol{W}(n) = \boldsymbol{d}_n^{'}$.

Substituting $\boldsymbol{R}(n) = \boldsymbol{C}_n^T \boldsymbol{C}_n$, $\boldsymbol{P}(n) = \tilde{\boldsymbol{X}}_n^T \boldsymbol{d}_n$, and the QR decomposition of $\boldsymbol{C}_n$ into the normal equation, one gets

$$\boldsymbol{C}_n^T \boldsymbol{C}_n \boldsymbol{W}(n) = \boldsymbol{P}(n) \Leftrightarrow \boldsymbol{R}_n^T \boldsymbol{Q}_n^T \boldsymbol{Q}_n \boldsymbol{R}_n \boldsymbol{W}(n) = \boldsymbol{R}_n^T \boldsymbol{Q}_n^T \tilde{\boldsymbol{d}}_n$$

Assuming $\boldsymbol{R}_n$ is nonsingular, we have $\boldsymbol{Q}_n \boldsymbol{R}_n \boldsymbol{W}(n) = \tilde{\boldsymbol{d}}_n$. Multiplying both sizes by $\boldsymbol{Q}_n^T$ and noting $\boldsymbol{Q}_n^T \boldsymbol{Q}_n = \boldsymbol{I}$, one gets

$$\boldsymbol{R}_n \boldsymbol{W}(n) = \boldsymbol{Q}_n^T \tilde{\boldsymbol{d}}_n = \boldsymbol{d}_n' \ .$$

b) **[6 marks]** In the LMS algorithm, the weight vector is updated in the negative direction of the gradient vector ∇:

$$\boldsymbol{W}_{n+1} = \boldsymbol{W}_n - \mu\nabla$$

where μ is a stepsize parameter, which is used to average out the effect of additive noise and the noise generated by approximating the gradient. The gradient is approximated by the instantaneous gradient by replacing ξ by $e(n)$, its instantaneous value:

$$\hat{\nabla} = \frac{\partial e^2(n)}{\partial \boldsymbol{W}} = -2e(n)\boldsymbol{X}.$$

Putting them together, one gets the LMS update as follows:

$$\boldsymbol{W}_{n+1} = \boldsymbol{W}_n + 2\mu e(n)\boldsymbol{X}.$$

It can be shown that the LMS algorithm converges in the mean when the stepsize satisfies

$$|1-\mu\lambda_k| < 1.$$

Hence, the range of values of μ that the LMS algorithm converges in the mean is

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

where $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{R}_{XX}$. Since $\boldsymbol{R}_{XX}$ is an autocorrelation matrix, its eigenvalues are *nonnegative*. Hence an upper bound on $\lambda_{\max}$ is

$$\lambda_{\max} < \sum_{k=0}^{M-1} \lambda_k = trace(\boldsymbol{R}_{XX}) = L \cdot r_{xx}(0),$$

where $r_{XX}(0)$ is the input signal power that is easily estimated, and $trace(\boldsymbol{A}) = \sum_{k=0}^{L-1} a_{kk}$ for any $(L \times L)$ matrix A. Therefore, an upper bound on the step size μ is

$$2/(L \cdot r_{xx}(0)).$$

The smaller the value of $|1-\mu\lambda_k|$, the faster is its convergence rate. Even if we choose the stepsize to be

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

The convergence rate of the LMS algorithm will however depend on the decay of the mode corresponding to the *smallest eigenvalue* $\lambda_{\min}$ at a rate

$$\left[\overline{W}(n)\right]_{\lambda_{\min}} \approx C\left(1-\frac{\lambda_{\min}}{\lambda_{\max}}\right)^{n} u(n),$$

where C is a constant and $u(n)$ is the unit step sequence.

Consequently, *the ratio* $(\lambda_{\min}/\lambda_{\max})$ *determines the convergence rate. If* $(\lambda_{\min}/\lambda_{\max})$ *is much smaller than unity, the convergence will be very slow. If* $(\lambda_{\min}/\lambda_{\max})$ *is close to unity, the convergence rate of the algorithm is fast.*

The total MSE at the output of the adaptive filter is

$$MSE = MSE_{\min} + J_{\mu},$$

where $MSE_{\min}$ is the minimum MSE , and J_{μ} is called the *excess error due to the noisy gradient estimate* (J_{Δ}) *and tracking errors* (J_l). The tracking errors result from the lag in track slowly time-variant signal statistics and it decreases with the stepsize, whereas the error due to the noisy graduate increases with increasing stepsizes. The optimal stepsize is obtained when the excess error due to the noisy gradient estimate is equal to that of the lag error.

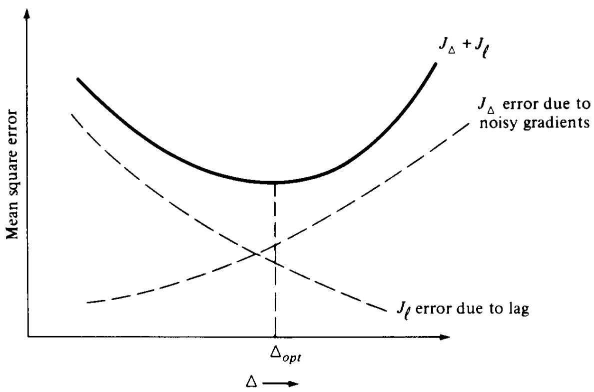


FIGURE 6.20 Excess mean-square error J_{Δ} and lag error J_l as a function of the step size Δ.

c)

i) **[5 marks]** The mean squared error (MSE) can be written as,

$$MSE = E[\varepsilon_n^2] = E[(x(n)-y(n))^2] = E[(x(n)-\sum_{k=1}^{P} a_i x(n-k))^2]$$

$$= E\left[x^2(n)\right] - 2\sum_{k=1}^{P} a_k E\left[x(n)x(n-k)\right] + \sum_{k=1}^{P} a_k \sum_{j=1}^{P} a_j E\left[x(n-k)x(n-j)\right]$$

$$= r_{xx}(0) - 2\boldsymbol{a}^T \boldsymbol{r}_{xx} + \boldsymbol{a}^T \boldsymbol{R}_{xx} \boldsymbol{a} .$$

ii) **[3 marks]** The gradient of the mean squared error function with respect to the filter coefficient vector is given by:

$$\frac{\partial}{\partial \boldsymbol{a}} E\left[e^2(n)\right] = -2\boldsymbol{r}_{xx}^T + 2\boldsymbol{a}^T \boldsymbol{R}_{xx} ,$$

where the gradient vector is defined as

$$\frac{\partial}{\partial \boldsymbol{a}} = \left[\frac{\partial}{\partial a_1}, \quad \frac{\partial}{\partial a_2}, \quad \frac{\partial}{\partial a_3}, \quad \ldots, \quad \frac{\partial}{\partial a_P} \right]^T .$$

The minimum is obtained by setting the gradient to zero as

$$\boldsymbol{R}_{xx}\boldsymbol{a} = \boldsymbol{r}_{xx} ,$$

or equivalently,

$$\boldsymbol{a} = (\boldsymbol{R}_{xx})^{-1} \boldsymbol{r}_{xx} .$$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & r_{xx}(2) & \cdots & r_{xx}(P-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(P-2) \\ r_{xx}(2) & r_{xx}(1) & & \cdots & r_{xx}(P-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}(P-1) & r_{xx}(P-2) & r_{xx}(P-3) & \cdots & r_{xx}(0) \end{bmatrix}^{-1} \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ \vdots \\ r_{xx}(P) \end{bmatrix}$$

iii) **[5 marks]**

[1 marks] By taking the z-transform of the linear prediction equation

$$e(n) = x(n) - \sum_{i=1}^{P} a_i x(n-i) ,$$

show that the transfer function from $x(n)$ to the prediction error $e(n)$ is

$$\frac{E(z)}{X(z)} = H(z) = 1 - \sum_{i=1}^{P} a_i z^{-i} .$$

Taking z-transform on both sides of (*) gives

$$E(z) = X(z) - \sum_{i=1}^{P} a_i z^{-i} X(z) = (1 - \sum_{i=1}^{P} a_i z^{-i}) X(z) = H(z) X(z) .$$

[3 marks] If $H(z) = 1 - \sum_{i=1}^{P} a_i z^{-i}$ has all its zeros ω_k, k=1,…,P, being distinct and are on the unit circle except $\omega_k = 0, \pi$ with $E(z) = e$ and ROC |z|>1, show that $x(n)$ can be written as a sum of P complex sinusoids as follows

$$x(n) = \sum_{k=1}^{P} A_k e^{j(n\omega_k)} u(n),$$

for some constant $A_k = |A_k| e^{j\phi_k}$. Moreover, if P is even and a_i's are real numbers, $x(n)$ can be written as

$$x(n) = [2\sum_{k=1}^{P/2} |A_k| \cos(\omega_k n + \phi_k)] \cdot u(n). \qquad (*)$$

If $H(z)$ has all its zero being distinct and on unit circle except $\omega_k = 0, \pi$, then it can be written as

$$H(z) = c\prod_{k=1}^{P} (z^{-1} - e^{j\omega_k}),$$

for some complex number c. Since $E(z) = e$, one gets

$$X(z) = \frac{e}{c\prod_{k=1}^{P} (z^{-1} - e^{j\omega_k})} = \sum_{k=1}^{P} \frac{A_k}{(z^{-1} - e^{j\omega_k})}, \ |z| > 1.$$

Taking inverse z-transform, one gets

$$x(n) = \sum_{k=1}^{P} A_k e^{jn\omega_k} u(n).$$

[1 marks] If P is even and a_i's are real numbers, thus the zeros come in complex conjugates and we have

$$x(n) = \sum_{k=1}^{P/2} [A_k e^{jn\omega_k} + A_k^* e^{-jn\omega_k}] u(n) = 2\sum_{k=1}^{P/2} |A_k| \cos(n\omega_k + \phi_k) u(n).$$

iv) **[1 marks]** Using the result in (iii) and assuming that the additive noise is small, suggest a method to estimate the frequencies of a multi-sinusoidal signal in form of (*) above.

By applying linear prediction to $x(n)$, we can obtain $H(z) = 1 - \sum_{i=1}^{P} a_i z^{-i}$. By finding its roots, we can identify the frequencies ω_k.

Question 5

a) **[4 marks]**

An estimator is unbiased if on the average the estimator will yield the true value of the unknown parameter. Mathematically,

$$E(\hat{\theta}) = \theta \ , \qquad a < \theta < b,$$

where θ is the true parameter, $\hat{\theta}$ is our estimate, and (a,b) is the range of possible values of θ.

$b(\hat{\theta}) = E(\hat{\theta}) - \theta$ is called the bias of the estimator.

b) **[2marks]**

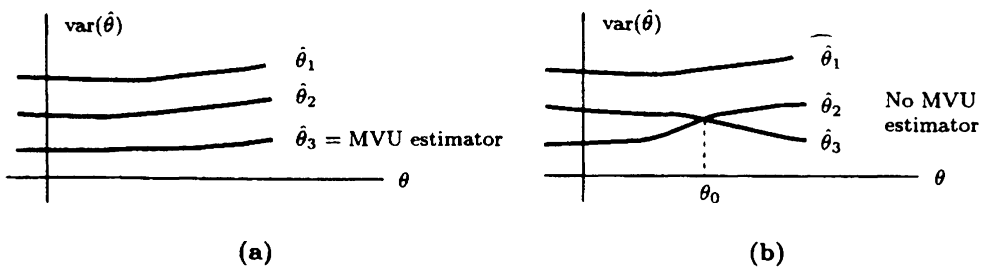
$$mse(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

It measures the average mean squared deviation of the estimator from the true value.

c) **[2 marks]**

In the minimum variance unbiased (MVU) estimator, the bias is constrained to be zero and the variance is minimized.

d) **[3 marks]**



Possible dependence of estimator variance with θ.

e)

i) **[3 marks]**

Write down the likelihood function $p(\boldsymbol{x};\boldsymbol{\theta})$ where $\boldsymbol{x} = (x[0],\ldots,x[N-1])^T$.

$$p(\boldsymbol{x};\boldsymbol{\theta} \mid \boldsymbol{\beta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))^2 \right\}.$$

ii) **[6 marks]**

Determine $\dfrac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta}\mid\boldsymbol{\beta})}{\partial\theta_i}$ and $\dfrac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta}\mid\boldsymbol{\beta})}{\partial\theta_i\theta_j}$.

$$\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta}) \ ,$$

and

$$\frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\theta_j} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}[-\nabla_{\theta_j} f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta}) + (x[n]-f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla^2_{\theta\theta_j} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})]$$

$$\frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}[-\nabla f_{\boldsymbol{\theta}}(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla f_{\boldsymbol{\theta}}(\boldsymbol{\beta}_n;\boldsymbol{\theta})^T + (x[n]-f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla^2 f_{\boldsymbol{\theta}}(\boldsymbol{\beta}_n;\boldsymbol{\theta})]$$

iii) **[4 marks]**

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\theta}) &= -E[\frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}] \\ &= \frac{1}{\sigma^2}E[\sum_{n=0}^{N-1}\{\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})^T - (x[n]-f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla^2_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})\}]\end{aligned}.$$

At the true parameter $x[n]-f(\boldsymbol{\beta}_n;\boldsymbol{\theta}) = w[n]$ (since the estimator is assumed to be unbiased) and hence we have

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\theta}) &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\{\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})^T - E[w[n]]\cdot\nabla^2_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})\} \\ &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})^T \ ,\end{aligned}$$

since $E[w[n]]=0$.

Alternatively, one can use the expression $\boldsymbol{I}(\boldsymbol{\theta}) = E[\dfrac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\dfrac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}^T]$ and get

$$\boldsymbol{I}(\boldsymbol{\theta}) = E[\left(\frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})\right)\left(\frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})\right)^T]$$

Using similar argument at the true parameter, one gets

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\theta}) &= \frac{1}{\sigma^4}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}E[w[n]w[m]]\cdot\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})\nabla_{\theta} f(\boldsymbol{\beta}_m;\boldsymbol{\theta})^T \ . \\ &= \frac{1}{\sigma^4}\sum_{m=0}^{N-1}\sigma^2\cdot\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})^T = \frac{1}{\sigma^2}\sum_{m=0}^{N-1}\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})\nabla_{\theta} f(\boldsymbol{\beta}_n;\boldsymbol{\theta})^T \ ,\end{aligned}$$

since $w[n]$ is white with variance σ^2.

For linear model, $f(\boldsymbol{\beta}_n;\boldsymbol{\theta}) = \boldsymbol{\beta}_n^T\boldsymbol{\theta}$. The corresponding FIM is

$$\boldsymbol{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\boldsymbol{\beta}_n \cdot (\boldsymbol{\beta}_n)^T .$$

iv) **[6 marks]** Determine the Cramer-Rao Lower Bound (CRLB) of $\boldsymbol{\theta}$.

Since $\boldsymbol{I}^{-1}(\boldsymbol{\theta}) = \sigma^2\left(\sum_{n=0}^{N-1}\nabla_\theta f(\boldsymbol{\beta}_n;\boldsymbol{\theta})\nabla_\theta f(\boldsymbol{\beta}_n;\boldsymbol{\theta})^T\right)^{-1}$, and $\boldsymbol{C}_\theta \geq \boldsymbol{I}^{-1}(\boldsymbol{\theta})$.

For linear estimation, we have $\boldsymbol{I}^{-1}(\boldsymbol{\theta}) = \sigma^2\left(\sum_{n=0}^{N-1}\boldsymbol{\beta}_n\boldsymbol{\beta}_n{}^T\right)^{-1}$ and the MLE is the MVU and hence achieve the CRLB.

For the linear fitting problem,

$$\boldsymbol{I}(\boldsymbol{\theta}) = -E\begin{bmatrix} \dfrac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial A^2} & \dfrac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial A \partial B} \\ \dfrac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial B \partial A} & \dfrac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial B^2} \end{bmatrix} = \frac{1}{\sigma^2}\begin{bmatrix} N & \dfrac{N(N-1)}{2} \\ \dfrac{N(N-1)}{2} & \dfrac{N(N-1)(2N-1)}{6} \end{bmatrix}$$

$$\boldsymbol{I}^{-1}(\boldsymbol{\theta}) = \sigma^2\begin{bmatrix} \dfrac{2(2N-1)\sigma^2}{N(N+1)} & -\dfrac{6}{N(N+1)} \\ -\dfrac{6}{N(N+1)} & \dfrac{12}{N(N^2-1)} \end{bmatrix}$$, and $\boldsymbol{C}_\theta - \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \geq 0$. Hence,

$$C_A \geq \frac{2(2N-1)\sigma^2}{N(N+1)},$$

$$C_B \geq \frac{12\sigma^2}{N(N^2-1)}.$$

v) **[4 marks]**

The MLE is obtained by minimizing the likelihood function, which is equivalent to the minimization of the nonlinear least squares function:

$$NLS(\boldsymbol{\theta}) = \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))^2 .$$

If the function is differentiable, one can set the partial derivatives to zero to obtain the following 1st order necessary condition for optimality:

$$\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n] - f(\boldsymbol{\beta}_n;\boldsymbol{\theta}))\cdot \nabla f_{\boldsymbol{\theta}}(\boldsymbol{\beta}_n;\boldsymbol{\theta}) = 0 .$$

For linear model, $f(\boldsymbol{\beta}_n;\boldsymbol{\theta}) = \boldsymbol{\beta}_n^T\boldsymbol{\theta}$, the corresponding MLE satisfies

$$\sum_{n=0}^{N-1}(x[n] - \boldsymbol{\beta}_n^T\boldsymbol{\theta})\cdot \boldsymbol{\beta}_n = \mathbf{0} \quad \Leftrightarrow \left(\sum_{n=0}^{N-1}\boldsymbol{\beta}_n\boldsymbol{\beta}_n^T\right)\boldsymbol{\theta} = \sum_{n=0}^{N-1}x[n]\boldsymbol{\beta}_n ,$$

which is the normal equation in linear estimation.

Question 6

a) **[4 marks]**

What is the main difference and possible advantage between the classical approach to statistical estimation such as maximum likelihood estimation and the Bayesian approach?

In the Bayesian approach, the parameter is treated as a random variable with a prior distribution. Therefore, prior knowledge of the parameters can be incorporated. This is impossible in conventional approach like the MLE. If appropriate prior knowledge can be employed, the estimation accuracy will be improved, especially when signal-to-noise ratio is large or when few measurements are available.

b) **[5 marks]** The Bayesian mean square error (Bmse) is defined as

$$Bmse(\hat{\boldsymbol{\theta}}) = \int\int \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2^2 p(\boldsymbol{x},\boldsymbol{\theta})d\boldsymbol{x}d\boldsymbol{\theta} ,$$

where the expectation is performed with respect to the joint PDF $p(\boldsymbol{x},\boldsymbol{\theta})$.

Using the Bayes' theorem $p(\boldsymbol{x},\boldsymbol{\theta}) = p(\boldsymbol{\theta} \mid \boldsymbol{x})p(\boldsymbol{x})$, we have

$$Bmse(\hat{\boldsymbol{\theta}}) = \int\int \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2^2 p(\boldsymbol{\theta} \mid \boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}d\boldsymbol{\theta}$$

$$= \int[\int \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2^2 \ p(\boldsymbol{\theta} \mid \boldsymbol{x})d\boldsymbol{\theta}]\cdot p(\boldsymbol{x})d\boldsymbol{x} .$$

Since $p(\boldsymbol{x})$ is positive, $Bmse(\hat{\boldsymbol{\theta}})$ is minimized when the term inside the bracket is minimized. Differentiating with respect to $\hat{\boldsymbol{\theta}}$, one gets

$$\frac{\partial}{\partial \hat{\boldsymbol{\theta}}}[\int \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2^2 \ p(\boldsymbol{\theta} \mid \boldsymbol{x})d\boldsymbol{\theta}] = -2\int(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})p(\boldsymbol{\theta} \mid \boldsymbol{x})d\boldsymbol{\theta} = 2\hat{\boldsymbol{\theta}} - 2\int \boldsymbol{\theta}\cdot p(\boldsymbol{\theta} \mid \boldsymbol{x})d\boldsymbol{\theta} .$$

Setting it to zero gives

$$\hat{\boldsymbol{\theta}} = \int \boldsymbol{\theta} \cdot p(\boldsymbol{\theta} \mid \boldsymbol{x}) d\boldsymbol{\theta} = E_{\theta}[\boldsymbol{\theta} \mid \boldsymbol{x}].$$

c)

Consider the state space model over time instants k=0,1,…

(6-1) $\boldsymbol{\theta}_k = \boldsymbol{A}_k \boldsymbol{\theta}_{k-1} + \boldsymbol{\varepsilon}_k$

(6-2) $\boldsymbol{x}_k = \boldsymbol{H}_k \boldsymbol{\theta}_k + \boldsymbol{w}_k$

where $\boldsymbol{\theta}_k$ is the mx1 state vector, $\boldsymbol{A}_k$ is the mxm state transition matrix, $\boldsymbol{\varepsilon}_k$ is a zero mean Gaussian distributed noise (commonly referred to as excitation/innovation) with covariance $\boldsymbol{C}_{\varepsilon_k}$, $\boldsymbol{x}_k$ is the dx1 measurement vector, $\boldsymbol{H}_k$ is the dxm measurement matrix, and $\boldsymbol{w}_k$ is the dx1 measurement noise which is $\boldsymbol{\theta}_k$, assumed to be Gaussian distributed with mean zero and covariance $\boldsymbol{C}_{w_k}$.

i) **[5 marks]** Show that the mean and variance of $\boldsymbol{\theta}_k$ given $\boldsymbol{\theta}_{k-1}$ and the dynamical equation (6-1) satisfy

$$\boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}} = E[\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}] = \boldsymbol{A}_k \boldsymbol{\mu}_{\theta_{k-1}}$$

and $$\boldsymbol{C}_{\theta_k \mid \theta_{k-1}} = E[(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k})((\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k})^T \mid \boldsymbol{\theta}_{k-1}] = \boldsymbol{A}_k \boldsymbol{C}_{\theta_{k-1}} \boldsymbol{A}_k^T + \boldsymbol{C}_{\varepsilon_k},$$

where $\boldsymbol{\mu}_{\theta_{k-1}}$ and $\boldsymbol{C}_{\theta_{k-1}}$ are the mean and covariance of $\boldsymbol{\theta}_{k-1}$. In other words, the dynamical equations allow us to predict the density of $\boldsymbol{\theta}_k$ given the previous one at $\boldsymbol{\theta}_{k-1}$.

ii) **[10 marks]** In (6-2), measurement $\boldsymbol{x}_k$ is taken so as to correct or update the density of $\boldsymbol{\theta}_k$ after the prediction from the dynamical equation. Let

$$\boldsymbol{z}_k = \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{y}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_k \boldsymbol{\theta}_k + \boldsymbol{w}_k \\ \boldsymbol{\theta}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_k & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_k \\ \boldsymbol{w}_k \end{bmatrix}.$$

As $\boldsymbol{z}_k$ is a linear transformation of a Gaussian vector $[\boldsymbol{\theta}_k^T, \boldsymbol{w}_k^T]^T$, it too is Gaussian distributed. Therefore, the mean and hence covariance (and hence the density) of $\boldsymbol{\theta}_k$ can again to updated using a classical result of Gaussian distribution relating its component vectors as $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$ as follows

(6-3) $E[\boldsymbol{y}_k \mid \boldsymbol{x}_k] = \boldsymbol{\mu}_{y_k} + \boldsymbol{C}_{y_k \boldsymbol{x}_k} \boldsymbol{C}^{-1}_{\boldsymbol{x}_k \boldsymbol{x}_k} (\boldsymbol{x}_k - E[\boldsymbol{x}_k])$

(6-4) $\boldsymbol{C}_{y_k \mid \boldsymbol{x}_k} = \boldsymbol{C}_{y_k} - \boldsymbol{C}_{y_k \boldsymbol{x}_k} \boldsymbol{C}^{-1}_{\boldsymbol{x}_k \boldsymbol{x}_k} \boldsymbol{C}_{\boldsymbol{x}_k y_k}$.

where $\boldsymbol{y}_k$ is now identified as $\boldsymbol{\theta}_k$. Due to the dynamical equation the mean and covariance of $\boldsymbol{y}_k$, i.e. $\boldsymbol{\theta}_k$, are given by $\boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}}$ and $\boldsymbol{C}_{\theta_k \mid \theta_{k-1}}$ as derived in part (i) above.

Using (6-2), show that

1) $E(\boldsymbol{x}_k) = \boldsymbol{H}_k E[\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}] = \boldsymbol{H}_k \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}}$.

2)
$$\begin{aligned}
\boldsymbol{C}_{\boldsymbol{x}_k} &= E[(\boldsymbol{x}_k - E(\boldsymbol{x}_k))(\boldsymbol{x}_k - E(\boldsymbol{x}_k)^T \mid \boldsymbol{\theta}_{k-1}] \\
&= E[(\boldsymbol{H}_k \boldsymbol{\theta}_k + \boldsymbol{w}_k - \boldsymbol{H}_k \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}})(\boldsymbol{H}_k \boldsymbol{\theta}_k + \boldsymbol{w}_k - \boldsymbol{H}_k \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}})^T] \\
&= E[\boldsymbol{H}_k (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}}) + \boldsymbol{w}_k)(\boldsymbol{H}_k (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}}) + \boldsymbol{w}_k)^T] \\
&= \boldsymbol{H} E[(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}})(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}})^T \boldsymbol{H}_k^T] + E[\boldsymbol{w}_k \boldsymbol{w}_k^T] \\
&= \boldsymbol{H}_k \boldsymbol{C}_{\theta_k \mid \theta_{k-1}} \boldsymbol{H}_k^T + \boldsymbol{C}_{w_k} ,
\end{aligned}$$

3)
$$\begin{aligned}
\boldsymbol{C}_{\theta_k \boldsymbol{x}_k} &= E[(\boldsymbol{\theta}_k - E(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}))(\boldsymbol{x}_k - E(\boldsymbol{x}_k))^T \mid \boldsymbol{\theta}_{k-1}] \\
&= E[(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}})(\boldsymbol{H}_k (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}}) + \boldsymbol{w}_k)^T] \\
&= E[(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}})(\boldsymbol{H}_k (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}}))^T] \\
&= \boldsymbol{C}_{\theta_k \mid \theta_{k-1}} \boldsymbol{H}_k^T .
\end{aligned}$$

iii) **[3 marks]** By substituting these results and those in part (i) into (6-3) and (6-4), show that

$$\boldsymbol{\mu}_{\theta_k} \equiv E(\boldsymbol{\theta}_k \mid \boldsymbol{x}_k) = \boldsymbol{\mu}_{\theta_k \mid \theta_{k-1}} + \boldsymbol{K}_k \boldsymbol{e}_k$$

and $\boldsymbol{C}_{\theta_k} \equiv \boldsymbol{C}_{\theta_k \mid x} = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k) \boldsymbol{C}_{\theta_k \mid \theta_{k-1}}$, where $\boldsymbol{K}_k = \boldsymbol{C}_{\theta_k|\theta_{k-1}} \boldsymbol{H}_k^T (\boldsymbol{H}_k \boldsymbol{C}_{\theta_k|\theta_{k-1}} \boldsymbol{H}_k^T + \boldsymbol{C}_{w_k})^{-1}$, $\boldsymbol{\mu}_{\theta_k|\theta_{k-1}} = \boldsymbol{A}_k \boldsymbol{\mu}_{\theta_{k-1}}$, $\boldsymbol{C}_{\theta_k|\theta_{k-1}} = \boldsymbol{A}_k \boldsymbol{C}_{\theta_{k-1}} \boldsymbol{A}_k^T + \boldsymbol{C}_{\varepsilon_k}$ and $\boldsymbol{e}_k = (\boldsymbol{x}_k - \boldsymbol{H}_k \boldsymbol{\mu}_{\theta_k|\theta_{k-1}})$ is the error vector in predicting the measurement $\boldsymbol{x}_k$ from the dynamical equation. The process can be repeated over time and is known as the Kalman filter.

iv) **[6 marks]** Consider the recursive estimation of the phase angle ϕ of a sinusoid embedded in zero-mean white Gaussian noise $w[n]$ with known variance σ_ε^2

$$x[k] = A\cos(2\pi f_0 k + \phi) + \varepsilon[n], \ k=0,1,2,\ldots$$

The amplitude A and frequency f_0 are assumed to be known.

By writing $\cos(2\pi f_0 k + \phi) = \cos(2\pi f_0 k)\cos(\phi) - \sin(2\pi f_0 k)\sin(\phi)$ and define the state vector $\boldsymbol{\theta}_k = [A\cos(\phi), A\sin(\phi)]^T$, illustrate how the Kalman filter above can be used to estimate the phase angle recursively.

[Hint: you can assume that $\boldsymbol{A}_k = \boldsymbol{I}$ and $\boldsymbol{C}_{w_k} = \sigma_w^2$ is known.]

Let $\boldsymbol{A}_k = \boldsymbol{I}$, $\boldsymbol{H}_k = \boldsymbol{X}_k^T = [\cos(2\pi f_0 k), -\sin(2\pi f_0 k)]$,

$$\boldsymbol{\mu}_{\theta_k} = \boldsymbol{\mu}_{\theta_k|\theta_{k-1}} + \boldsymbol{K}_k \boldsymbol{e}_k$$

and $\boldsymbol{C}_{\theta_k} \equiv \boldsymbol{C}_{\theta_k|x} = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{X}_k^T)\boldsymbol{C}_{\theta_k|\theta_{k-1}}$,

where $\boldsymbol{K}_k = \dfrac{\boldsymbol{C}_{\theta_k|\theta_{k-1}} \boldsymbol{X}_k}{\sigma_w^2 + \boldsymbol{X}_k^T \boldsymbol{C}_{\theta_k|\theta_{k-1}} \boldsymbol{X}_k}$, $\boldsymbol{\mu}_{\theta_k|\theta_{k-1}} = \boldsymbol{\mu}_{\theta_{k-1}}$, $\boldsymbol{C}_{\theta_k|\theta_{k-1}} = \boldsymbol{C}_{\theta_{k-1}} + \boldsymbol{C}_{\varepsilon_k}$ and $\boldsymbol{e}_k = (\boldsymbol{x}_k - \boldsymbol{X}_k^T \boldsymbol{\mu}_{\theta_k|\theta_{k-1}})$ is the error vector in predicting the measurement $\boldsymbol{x}_k$ from the state dynamic.

From $\boldsymbol{\theta}_k = [Ac_k, As_k] \approx [A\cos(\phi_k), A\sin(\phi_k)]^T$, one can estimate ϕ_k from $\phi_k = \tan^{-1}(As_k / Ac_k)$.