EE4-13
EE9-SO15

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2013

MSc and EEE/EIE PART IV: MEng and ACGI

Corrected Copy

SPECTRAL ESTIMATION AND ADAPTIVE SIGNAL PROCESSING

Friday, 17 May 10:00 am

Time allowed: 3:00 hours

There are FIVE questions on this paper.

Answer TWO of questions 1,2,3 and ONE of questions 4,5.

All questions carry equal marks

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : D.P. Mandic, D.P. Mandic
Second Marker(s) : M.K. Gurcan, M.K. Gurcan

1) Nonparametric spectrum estimation relies on accurate estimation of the covariance of data under consideration.

a) State the expressions for the bias and variance of the periodogram $P_{per}(\omega)$, and explain the physical meaning behind the bias B and variance var in spectrum estimation. [5]

b) The fast Fourier transform (FFT) is most effective if the length of time series N is a power of 2. To ensure this, we might pad the time series with zeros until $N_{zp} = 2^m$, i.e. [5]

$$x_{zp}(n) = \begin{cases} x(n), & 0 \leq n < N-1 \\ 0, & N \leq n < N_{zp} - 1 \end{cases}$$

Comment on the bias and variance of a zero-padded periodogram, and on the advantages and disadvantages of zero padding.

c) The Blackman-Tukey spectral estimator is given by [5]

$$\hat{P}_{BT}(\omega) = \sum_{k=-M}^{M} \hat{r}_x(k) w(k) e^{-\jmath k\omega}$$

where $\hat{r}_x$ denotes the autocorrelation function (ACF) estimate and $w(k)$ is the lag window applied to the ACF estimate. Let $P(\omega)$ denote the true power spectral density. Show that the values of the time lag window $w(k)$ that minimise the squared bias

$$B^2(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[P(\omega) - E\{\hat{P}_{BT}(\omega)\} \right]^2 d\omega$$

correspond to those of a rectangular window, and comment on the relation between the bias and spectral resolution. The variables with a circumflex are estimates, e.g. $\hat{\sigma}^2$ is an estimate of the variance σ^2.

$\Big($ Hint: Without loss in generality assume $E\{\hat{\mathbf{r}}_x(k)\} = \alpha(k)\mathbf{r}(k)$. Also employ Parseval's equality, $\frac{1}{2\pi}\int_{-\pi}^{\pi} |P(\omega))|^2 d\omega = \sum_{k=-\infty}^{\infty} |r(k)|^2$ $\Big)$

d) Performance of nonparametric spectral estimators is often expressed in terms of [5]

$$\text{variability of the estimate} \qquad \nu = \frac{var\left\{\hat{P}_x(\omega)\right\}}{E^2\left\{\hat{P}_x(\omega)\right\}}$$

$$\text{figure of merit} \qquad \mathcal{M} = \nu \times \Delta\omega$$

Comment on the physical meaning of these performance criteria and explain in your own words the difference in performance between the periodogram and the Blackman-Tukey method, in view of these criteria. (Hint: the resolution of periodogram is $0.89 \times 2\pi/N$)

2) Several spectral estimation methods employ truncation or extrapolation of an estimated autocovariance sequence (ACS).

a) Suppose that $\{r(k)\}_{k=-\infty}^{\infty}$ is a valid autocovariance sequence, in other words, the ACS satisfies $\sum_{k=-\infty}^{\infty} r(k)e^{-\jmath\omega k} \geq 0$, for all ω, and assume that $r(k) = \alpha^{|k|}$, where $0 < \alpha < 1$. Establish whether it is possible that for some integer p, the truncated ACS

$$\sum_{k=-p}^{p} r(k)e^{-\jmath\omega k}$$

produces a negative power spectrum for some ω. Elaborate in more detail for $p = 1$ and justify your answer. [5]

b) For a data sequence $\{x(n)\}$ of length N, two frequently used spectral estimators are given by

$$\text{Bartlett:} \quad \hat{P}_B(\omega) = \frac{1}{K}\sum_{m=0}^{K-1} \hat{P}_{per}^{(m)}(\omega) \quad \text{where} \quad \hat{P}_{per}^{(m)} = \frac{1}{L}\Big|\sum_{n=0}^{L-1} x_m(n)e^{-\jmath n\omega}\Big|^2 \quad \text{and} \quad KL = N$$

$$\text{Blackman-Tukey:} \quad \hat{P}_{BT}(\omega) = \sum_{k=-M}^{M} \hat{r}_x(k)w(k)e^{-\jmath k\omega} \qquad M < N$$

i) Establish a functional relationship between the Bartlett and the Blackman-Tukey method, where the latter uses a rectangular window. Based on this relationship, explain which of the two is expected to yield higher variance and state the reasons for that. [5]

ii) Establish a functional relationship between the Welch method (averaging windowed periodograms) and the Blackman-Tukey method. Elaborate on the similarities and differences between the Welch and Blacmkan-Tukey methods for the 50% overlap of data segments, and comment on their accuracies and computational complexities.
(Hint: We employ a symmetric positive definite window sequence $w(n)$. To this end, $w(n)$ should be an auto-convolution of another window sequence $v(n)$, that is $w(n) = v(n) * v(n)$.) [5]

c) Now consider the maximum entropy spectral estimator. Explain the advantages and disadvantages of this approach, and the effects that the extrapolation of the autocovariance function within this method has on the shape of the so obtained spectral estimate. Provide your own justification and sketch the reasoning behind this approach, there is no need for a full derivation of the method. [5]

3) Accurate estimation of the autocovariance matrix is a pre-requisite for unbiased and minimum variance spectral estimation.

a) For a data sequence $x[0], x[1], \ldots, x[N-1]$ of length N, the *standard unbiased* estimate, $\hat{r}_{su}(k)$ and the *standard biased* estimate $\hat{r}_{sb}(k)$ of the true autocovariance function (ACF), $r_x(k)$ are respectively given by [5]

$$\begin{aligned}\hat{r}_{su}(k) &= \frac{1}{N-k}\sum_{n=0}^{N-1-k} x(n+k)x(n) = \frac{N-k}{N} r_x(k) \\ \hat{r}_{sb}(k) &= = \frac{1}{N} x_N(k) * x_N(-k) = \frac{1}{N}\sum_{n=0}^{N-1-k} x(n+k)x(n)\end{aligned}$$

Comment on the accuracy of these estimators for large lags k, and whether there is any range of k where the two estimators produce similar results. Does any of the above estimators produce a positive semidefinite sequence $\hat{r}_x$?

b) If the statistics of $x(n)$ are unknown, then the autocovariance matrix $\mathbf{R}_{xx}$ is also unknown and its inverse $\mathbf{R}_{xx}^{-1}$ cannot be evaluated. However, suppose we approximate $\mathbf{R}_{xx} = E\{\mathbf{x}(n)\mathbf{x}^T(n)\}$ at time n as follows [5]

$$\hat{\mathbf{R}}_{xx}(n) = \mathbf{x}(n)\mathbf{x}^T(n)$$

and use as the n–th order approximation to $\mathbf{R}_{xx}^{-1}$

$$\hat{\mathbf{R}}_{xx}^{-1}(n) = \mu \sum_{k=0}^{n} \left[\mathbf{I} - \mu \mathbf{x}(k)\mathbf{x}^T(k)\right]^k$$

Express $\hat{\mathbf{R}}_{xx}^{-1}(n+1)$ in terms of $\hat{\mathbf{R}}_{xx}^{-1}(n)$ and use this expression to derive a recursion for the weight vector $\mathbf{w}_{n+1}$ within the Wiener filter.

c) Suppose that a radar is tracking an object which is moving in a radial direction away from the radar, at a constant velocity v. The signal emitted by the radar is $Ae^{\jmath\omega n}$. Show that the backscattered signal, received by the radar after reflection off an object, is given by [5]

$$s(n) = Be^{\jmath(\omega-\omega_d)n} + w(n)$$

where $w(n)$ is the measurement noise, $\omega_d = 2\omega v/c$ is the so called Doppler frequency, and $B = \mu A e^{-2\jmath\omega r/c}$, where c is the speed of the radar wave, r is the distance to the object, and μ is the attenuation coefficient. Show that the speed measurement using a Doppler radar can be cast into a frequency estimation problem.

d) Explain in your own words how would you perform frequency estimation of q sinusoids, where p out of the q sinusoids are known. Give some intuition about the properties of the correlation matrices involved and their rank. [5]

4) Consider adaptive mean-square error (MSE) estimation. Assume that for a filter of length L, the teaching signal can be expressed as

$$d(n) = \mathbf{x}^T(n)\mathbf{w}_{opt} + q(n), \qquad q \sim \mathcal{N}(0, \sigma_q^2), \quad n = 0, \ldots, N-1$$

a) For the error surface shown in Figure 4.1, answer the following.

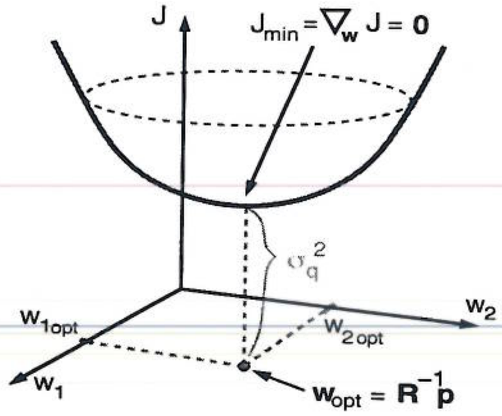


Fig. 4.1. Error surface of an $L = 2$ taps long adaptive filter

i) Show that the MSE as a function of filter weights $\mathbf{w}$ can be expressed as [4]

$$J(\mathbf{w}) = E\{|e(n)|^2\} = \sigma_d^2 - 2\mathbf{w}^T\mathbf{p} + \mathbf{w}^T\mathbf{R}\mathbf{w}$$

where $e(n) = d(n) - y(n) = d(n) - \mathbf{w}^T\mathbf{x}(n)$.

ii) Show that the minimum achievable MSE is $J_{min} = \sigma_q^2$, for $\mathbf{w} = \mathbf{w}_{opt}$. Next, explain the notion of a learning curve and illustrate their behaviour by sketching the learning curves for three different values of noise q, denoted by $q1$, $q2$ and $q3$, where $\sigma_{q1}^2 > \sigma_{q2}^2 > \sigma_{q3}^2$. [4]

iii) Explain the notions of excess mean-square error, misadjustment, and misalignment. It is well known that the misadjustment $\mathcal{M} \approx \mu L \sigma_x^2$. Explain in your words how the values of the stepsize μ, filter length L, and the input signal power σ_x^2 affect the misadjustment. [4]

b) We wish to design an adaptive predictor for a real-valued process $x(n)$ which is observed using noisy measurements $y(n)$ given by

$$y(n) = x(n) + q(n), \qquad q(n) \sim \mathcal{N}(0, \sigma_q^2), \qquad q \text{ is orthogonal to } x \ \ (q \perp x)$$

i) Using the least mean square (LMS) algorithm given by [4]

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n) y(n)$$

find the range of values for μ for which the LMS algorithm converges in the mean, and find $\lim_{n\to\infty} E\{\mathbf{w}(n)\}$.
(Hint: The i-th eigenvalues for y and x, denoted by λ_i^y and λ_i^x, are related as $\lambda_i^y = \lambda_i^x + \sigma_q^2$)

ii) The γ-LMS algorithm is capable of combating the effect of measurement noise. For the noisy observations $y(n)$, the γ-LMS is given by [4]

$$\mathbf{w}(n+1) = \gamma\mathbf{w}(n) + \mu e(n)\mathbf{y}(n), \qquad \gamma = \text{constant}$$

Explain how the γ-LMS algorithm can be used to remove the bias in the steady state of the LMS. Specifically, how would you select the values for μ and γ?

5) Adaptive filters can be connected to the environment in different ways (configurations), while for the same configuration they can be trained using different algorithms.

a) When used within the inverse system modelling configuration, adaptive filters can adaptively estimate an unknown time-varying transfer function.

i) Sketch the block diagram of an adaptive inverse system modelling configuration, explain its operation, and identify some applications where this scheme would be useful. [4]

ii) Sketch the block diagram of a single-weight adaptive filter, that is, a filter whose length is $L = 1$. Write a least mean square (LMS) update equation for this filter, and find the expression for the error function for a constant input $\mathbf{x}(n) = K$. [4]

iii) For a constant input $x(n) = K$ to the filter in ii), find the system function $H(z) = \frac{D(z)}{E(z)}$ relating $e(n)$ and $d(n)$, and determine the range of values for the stepsize parameter μ for which the so obtained $H(z)$ is stable. [4]

(Hint: it is easier to obtain the result when relating $e(n) - e(n-1)$ to $d(n) - d(n-1)$.)

b) Consider a modified cost function for the LMS-type adaptation, given by

$$J'(n) = \frac{1}{2}\, e^2(n) + \frac{\beta}{2}\, \mathbf{w}^T(n)\mathbf{w}(n)$$

where $\beta > 0$, $\mathbf{w}(n)$ is the filter weight vector, and $(\cdot)^T$ the vector transpose operator.

i) The cost function $J'(n)$ has two terms, one minimising the mean square error and the second penalising for large values of the norm of the weight vector. Explain in your own words the principle behind such a cost function. Derive the LMS coefficient update equation for $\mathbf{w}(n)$ that minimises $J'(n)$. [4]

ii) If the stepsize μ is small enough so that $\mathbf{w}(n)$ converges in the mean, find the expression for $\mathbf{w}(\infty)$. In other words, what does it converge to when $n \to \infty$? [4]

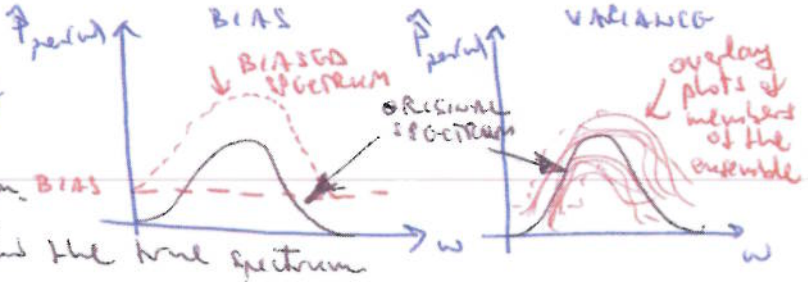
(Q1) [bookwork and own reasoning]

(a) Let $P_x(\omega)$ be the true power spectrum, and $\hat{P}_{per}(\omega)$ the periodogram estimate.
Then:
BIAS: $B = E\{\hat{P}_{per}(\omega)\} - P_x(\omega)$

VARIANCE: $var(\hat{P}_{per}(\omega)) = E\{[\hat{P}_{per}(\omega) - E\{\hat{P}_{per}(\omega)\}]^2\}$

Physical meaning:
A biased periodogram adds a spurious DC component to the spectrum, across all f. In other words, it adds white noise to the spectrum.

we desire: $\lim_{N\to\infty} E\{\hat{P}_{per}(\omega)\} = P_x(\omega)$ ← asymptotically unbiased



The variance corresponds to the deviation, in the mean square sense, from the true spectrum. It is reflected in the "thickness" of the overlay plots (right hand plot) around the true spectrum.

(b) by zero padding, $x_{zp}(n) = [\underbrace{x(0), \ldots, x(N-1), 0, \ldots, 0}_{2^m}]$ [new example]

- since we now have $2^{N_{zp}}$ points in DFT, and $N_{zp} > N$, the spectrum can reveal valleys and peaks not visible with N points.
- If an estimator is already asymptotically unbiased, zero padding will not affect it,

$$E\{\hat{P}_{per}(\omega)\} = \sum_{k=-(N-1)}^{N-1} E\{\hat{r}_{xx}(k)\} e^{-j2\pi fk} = \sum_{k=-(N-1)}^{N-1} \frac{N-|k|}{N} r_{xx}(k) e^{-j2\pi fk}$$ ← Bartlett window

The variance of the periodogram $var\{\hat{P}_{per}(f)\} = P_{xx}^2(f)\left[1 + \left(\frac{\sin 2\pi N f}{N \sin 2\pi f}\right)^2\right] \sim P_x^2(f)$

⇒ The variance will not be affected by zero padding (in terms of order of magnitude), although a longer data segment N_{zp} will reduce the effects of the 2^{nd} term in brackets.

ADVANTAGES: reveals details in true spectrum not visible with N points

DISADVANTAGES: no ~~insight~~ reduction of variance or increase in resolution.

(c)
$$E\{\hat{P}_{BT}(\omega)\} = \sum_{k=-(M-1)}^{M-1} w(k) r(k) e^{-j\omega k} = \sum_{k=-\infty}^{\infty} w(k) r(k) e^{-j\omega k}$$ FOR THE B-T method → $w(k)=0, |k| \geq M$

Using Parseval's equality:
$$B^2 = \sum_{k=-\infty}^{\infty} |r(k) - w(k)r(k)|^2 = \sum_{|k|\geq M} |r(k)|^2 + \sum_{|k|<M} (1-w(k))^2 |r(k)|^2$$

⇒ B^2 is minimised for $w(k)=1$, that is for a rectangular window. This is expected, as for good resolution we need a narrow main lobe of the spectral window function.

(d) The variability of the estimate V is effectively the normalised variance of the estimator. Then, for the periodogram, since $\boxed{var(\hat{P}_{per}(\omega)) \approx P^2(\omega), \Rightarrow V = 1}$

The Blackman-Tukey method:
$$\hat{P}_{BT}(\omega) = \sum_{k=-M}^{M} \hat{r}_x(k) e^{-j\omega k}$$ → $\boxed{\text{resolution} \sim \frac{1}{M}, \text{ variance} \sim M/N, \Rightarrow V_{BT} \sim \frac{M}{N}}$

The figure of merit:
$\mathcal{M} = V \times \Delta\omega$ (product of variability & resolution)

For $\hat{P}_{per}(\omega)$, $V=1$ and resolution $0.89\frac{2\pi}{N} \Rightarrow \mathcal{M} = 0.89\frac{2\pi}{N}$

For $\hat{P}_{BT}(\omega)$, $V \approx \frac{M}{N}$, resolution $\Delta\omega = 0.64\frac{2\pi}{N} \Rightarrow \mathcal{M} = 0.43\frac{2\pi}{N}$

Q2 [a), b) - new examples, c) book work]

a) We should be careful with truncation, as a truncated sum may yield a negative-definite correlation matrix.

For $r(k) = \alpha^{|k|}$ $\rightarrow$ $P(\omega)$ is a sum of cosines $P(\omega) = \sum_{k=-p}^{p} \alpha^{|k|} e^{-j\omega k} = 1 + 2\alpha\cos\omega + 2|\alpha|^2\cos 2\omega + \dots$

$\rightarrow$ $\boxed{\text{cosines may assume both positive and negative values.}}$

For $p=1$, $\sum_{k=-1}^{1} r(k) e^{-j\omega k} = \sum_{k=-1}^{1} \alpha^{|k|} e^{-j\omega k} = 1 + \cos 2\omega$ $\rightarrow$ $\boxed{\text{negative for } \omega = \pi}$

b) The Bartlett spectral estimate $\hat{P}_B(\omega) = \frac{1}{K}\sum_{m=0}^{K-1} \hat{P}_{per}^{(m)}(\omega)$ ... (1) $\quad \hat{P}_{BT}(\omega) = \sum_{k=-(M-1)}^{M-1} \hat{r}_x(k) w(k) e^{-jk\omega}$ (2)

$\hat{P}_B$ $\rightarrow$ data length for each $\hat{P}^{(m)}$ is L $(K \cdot L = N)$ and the resolution for every $\hat{P}^{(m)}(\omega)$ is $\frac{1}{L}$.
However the variance is reduced by factor K.

Now: $\hat{P}_{B}(\omega) = \sum_{k=-(M-1)}^{M-1} \left[\frac{1}{K}\sum_{n=0}^{K-1} \hat{r}_j(k) \right] e^{-j\omega k} = \sum_{k=-(M-1)}^{M-1} \hat{r}_B(k) e^{-j\omega k}$

which is obtained by inserting $\hat{P}_{per}^{(m)}(\omega) = \frac{1}{L}\left| \sum_{n=0}^{L-1} x(n) e^{-j\omega n} \right|^2 = \sum_{k=-L-1}^{L-1} \hat{r}_m(k) e^{-j\omega k}$ into (1)

$\Rightarrow$ $\hat{P}_B(\omega)$ is similar in form to the B-T estimator that uses a rectangular window. However, $\hat{r}_B(k)$ does not make efficient use of the available lag products, especially for $|k|$ near $(M-1)$. $\Rightarrow$ The variance of $\hat{P}_B(\omega)$ will be higher than when using the BT method. $\hat{P}_B$ uses a rectangular window $\Rightarrow$ less flexibility than $\hat{P}_{BT}$.

c) Similarly to the above:

$$\hat{P}_W(\omega) = \frac{1}{KLU}\sum_{i=0}^{K-1}\left| \sum_{n=0}^{L-1} w(n) x(n+iD) e^{-j\omega n} \right|^2$$

Consider 50% window overlap $\Rightarrow$ $\hat{P}_{per}^{(i)}(\omega) = \frac{1}{LU}\left| \sum_{n=0}^{L-1} v(n) x^{(i)}(n) e^{-j\omega n} \right|^2$

where U denotes the "power" of a temporal window $v(t)$, that is $U = \frac{1}{L}\sum_{n=0}^{L-1} |v(n)|^2$ $\Rightarrow$ $\hat{P}_W(\omega) = \frac{1}{K}\sum_{k=1}^{K} \hat{P}_{per}^{(i)}(\omega)$

Now: $\hat{P}_W(\omega) = \frac{1}{K}\sum_{k=1}^{K} \frac{1}{LU}\sum_{n=0}^{L-1}\sum_{m=0}^{L-1} v(n) v(m) x^{(i)}(n) x^{(i)}(m) e^{-j\omega(m-n)}$

$$= \frac{1}{LU}\sum_{n=0}^{L-1}\sum_{m=0}^{L-1} v(n) v(m) \left[\frac{1}{K}\sum_{k=1}^{K} x^{(i)}(n) x^{(i)}(m) \right] e^{-j\omega(n-m)}$$

For large N, the ACF estimate is sufficiently good, and

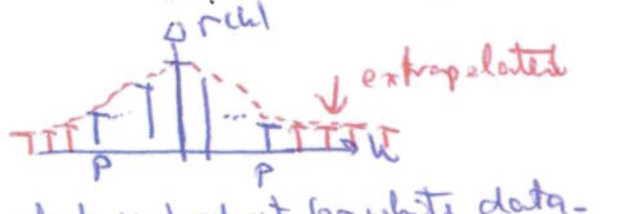
$\hat{P}_W(\omega) \approx \frac{1}{LU}\sum_{n=0}^{L-1}\sum_{m=0}^{L-1} v(n) v(m) \hat{r}(n-m) e^{-j\omega(n-m)}$ $\Rightarrow$ $\hat{P}_W(\omega) = \sum_{\tau=-(L-1)}^{L-1} w(\tau) \hat{r}(\tau) e^{-j\omega\tau}$ $\quad w(\tau) = v(n) * v(n)$

which is a form of the B-T estimator.
The Welch estimator can be computed through FFT. The B-T is theoretically preferred.

d) The MEM extrapolates the ACF so that $P_x(\omega) = \sum_{k=-p}^{p} r_x(k) e^{-j\omega k} + \sum_{|k|>p} r_e(k) e^{-j\omega k}$

The MEM does it in an "as flat as possible" way.

This, in turn, imposes an all-pole model on the data.



This is expected, since MEM maximizes spectral entropy $H = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln P_{xx}(\omega) d\omega$, which is highest for white data.

Q3 [a) bookwork and reasoning, b), c), d) - new examples]

a) * The standard unbiased estimate

$$E\{\hat{r}_x(k)\} = \frac{1}{N}\sum_{n=0}^{N-1-k} E\{x(n+k)x^*(n)\} = \frac{1}{N-k}\sum_{n=0}^{N-1-k} x(n+k)\,x(n) = \frac{N-k}{N}\, r_x(k)$$

- This estimate can take erratic values for large values of k
- it is not guaranteed to give a positive semidefinite $\hat{r}_x(k)$

* THE STANDARD BIASED ESTIMATE: $\hat{r}_x(k) = \sum_{n=-\infty}^{\infty} x_N(n+k)\, x_N^*(k) = \frac{1}{N} x_N(k) * x_N(k) =$

this estimate is biased

but it produces positive semidefinite $\hat{r}_x(k)$ $\quad = \frac{1}{N}\sum_{n=0}^{N-1-k} x(n+k)\,x(n)$

For small k, the BIASED and UNBIASED estimators are similar

b) Using the approximation $\hat{\underline{R}}_{xx} = \underline{x}(n)\underline{x}^T(n)$, we have

$$\underline{R}^{-1}_{xx}(n+1) = [\underline{I} - \mu\,\underline{x}(n)\,\underline{x}^T(n)]\,\underline{R}^{-1}_{xx}(n) + \mu\,\underline{I}$$

Multiplying both sides with $\underline{r}_{xd}$ from the right, we have

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

c) The received signal is an attenuated and delayed replica of the transmitted signal. The delay time = round trip from radar to object,

$$\text{delay} = 2\,\frac{r + v t}{c}$$

$\Rightarrow$ The reflected signal measured by the radar is

$$s(n) = \mu A\, e^{j\omega(n - 2\frac{r + v n}{c})} + w(n)$$
$$= \mu A\, e^{-j\omega 2r/c}\, e^{j(\omega - 2\omega v/c)n} + w(n)$$
$$= B\, e^{j(\omega - \omega_d)n} + w(n) \qquad \omega_d = \text{Doppler frequency}$$

Since ω and c are known, the estimation of v boils down to the determination of frequency of the sinusoid $s(n)$. Also, the amplitude B can be used to determine the range, if we know μ.

d) For p out of the q sinusoids known we can write

$$y(n) = \sum_{i=1}^{p} \hat{A}_i\, e^{j(\hat{\omega}_i n + \hat{\phi}_i)} + \sum_{i=1}^{q-p} A_i\, e^{j(\omega_i n + \phi_i)} + w(n)$$

$$= \hat{x}(n) + x(n) + w(n)$$

~~The problem boils down to estimating $\hat{x}(n)$ (unknown sinusoids)~~

We can now use any of the line spectra methods to estimate the frequencies of the unknown sinusoids, while the "known" sinusoids merge with noise. This is not straightforward as the ACF of "noise" will not be white and eigenmethods can only be applied after separating the eigenvalues of noise and eigenvalues of known sinusoids.

In this particular case, AR spectrum estimation may be a better choice.

Q4 [a) bookwork and intuition, b) new example] S4

i) We know that $e(n) = d(n) - \underline{w}^T(n)\,\underline{x}(n)$ and so

$$E\{|e(n)|^2\} = E\{e(n)e^T(n)\} = E\{(d(n) - \underline{w}^T(n)\underline{x}(n))(d(n) - \underline{w}^T(n)\underline{x}(n))^T\} =$$

$$= E\{d^2(n) - 2\,\underline{w}^T(n)\,d(n)\,\underline{x}(n) + \underline{w}^T(n)\,\underline{x}(n)\,\underline{x}^T(n)\,\underline{x}(n)\}$$

Since $\underline{w}$ are coefficients, we have

$$J(\underline{w}) = \underbrace{E\{d^2(n)\}}_{\sigma_d^2} - 2\,\underline{w}^T \underbrace{E\{d(n)\underline{x}(n)\}}_{\underline{r}_{dx} = \underline{p}} + \underline{w}^T \underbrace{E\{\underline{x}(n)\underline{x}^T(n)\}}_{\underline{R}}\,\underline{w}$$

$$\Rightarrow J(\underline{w}) = \sigma_d^2 - 2\underline{w}^T\underline{p} + \underline{w}^T\underline{R}\,\underline{w} \quad (1)$$

ii) For $d(n) = \underline{x}^T(n)\,\underline{w}_{opt} + g(n)$, replace in (1) to obtain

$\underline{w}_{opt} = \underline{R}^{-1}\underline{p}$

~~$J(\underline{w}) = \sigma_d^2 - 2\,\underline{w}^T E\{$~~

$$J(\underline{w}_{opt}) = E\{\underline{w}_{opt}^T\underline{x}\,d + g\,d\} - 2\underline{w}_{opt}^T\underline{p} + \underline{w}_{opt}\underbrace{\underline{R}\,\underline{R}^{-1}}_{\underline{I}}\underline{p}$$

$$= \underline{w}_{opt}^T\underline{p} + \sigma_g^2 - 2\,\underline{w}_{opt}^T\underline{p} + \underline{w}_{opt}\,\underline{p}$$

$$= \sigma_g^2$$ $\rightarrow$ exactly as shown on the error surface

iii)

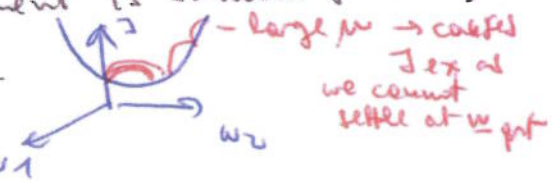
$$MSE(n) = J(n) = J_{opt} + J_{ex}$$

J_{ex}: excess MSE due to wrong filter length, large μ or some other design factor

MISALIGNMENT $\underline{V}(n) = \underline{w}_{opt} - \underline{w}(n)$

can be used as a performance criterion

$M \approx \mu L \sigma_x^2$, that is the misadjustment is smallest for small μ and small filter length



b) i) The LMS: $\underline{w}(n+1) = \underline{w}(n) + \mu\,e(n)\,\underline{y}(n)$

$\rightarrow$ ~~$E\{\underline{w}(n+1)\} = [\underline{I} - \mu\underline{R}_y]\,E\{\underline{w}(n)\}$~~

$$E\{\underline{w}(n+1)\} = [\underline{I} - \mu\,\underline{R}_y]\,E\{\underline{w}(n)\} + \mu\,\underline{r}_{dy}$$

Since $\underline{R}_y = \underline{R}_x + \sigma_g^2\underline{I} \Rightarrow \lambda_i^y = \lambda_i^x + \lambda_i^g = \lambda_i^x + \sigma_g^2$

$$\Rightarrow 0 < \mu < \frac{2}{\lambda_{max} + \sigma_g^2} \quad \text{and} \quad \lim_{n\to\infty} E\{\underline{w}(n)\} = [\underline{R}_x + \sigma_g^2\underline{I}]^{-1}\underline{r}_{dx}$$

ii)

$$\underline{w}(n+1) = \gamma\,\underline{w}(n) + \mu\,e(n)\,\underline{y}(n) = \gamma\,\underline{w}(n) + \mu\,[d(n) - \underline{w}^T(n)\,\underline{y}(n)]\,\underline{y}(n)$$

$$= [\gamma\underline{I} - \mu\,\underline{y}(n)\,\underline{y}^T(n)]\,\underline{w}(n) + \mu\,d(n)\,\underline{y}(n)$$

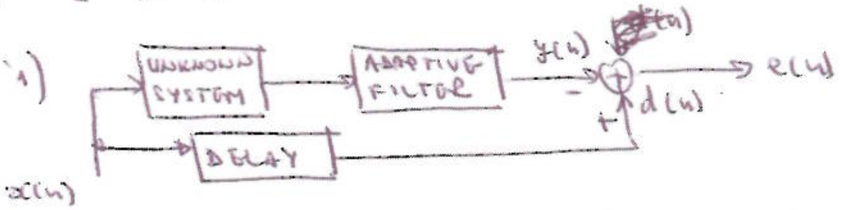
$\rightarrow$ taking the expectation and using the independence assumption:

$$E\{\underline{w}(n+1)\} = [\gamma\underline{I} - \mu(\underline{R}_x + \sigma_g^2\underline{I})]\,E\{\underline{w}(n)\} + \mu\,\underline{r}_{dy}$$

$$= [(\gamma - \mu\sigma_g^2)\,\underline{I} - \mu\,\underline{R}_x]\,E\{\underline{w}(n)\} + \mu\,\underline{r}_{dy}$$

$\rightarrow$ the conditions on μ and γ are $\gamma = 1 + \mu\sigma_g^2$, $0 < \mu < 2/\lambda^x_{max}$

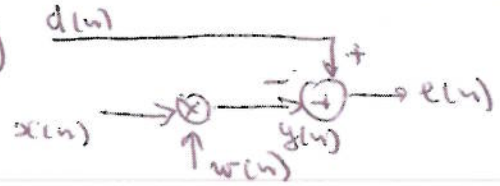
a) i)



This AF aims to find the inverse of the transfer function of the unknown system in real time.

APPLICATIONS: Channel equalisation in communications

ii)



← single weight adaptive filter

The LMS update for $N=1$ is

$w(n+1) = w(n) + \mu e(n) x(n)$

For $x(n) = K = \text{const} \Rightarrow e(n) = d(n) - K w(n)$ & $w(n) = w(n-1) + \mu K e(n-1)$

$\Rightarrow e(n) - e(n-1) = d(n) - d(n-1) - K[w(n) - w(n-1)] = d(n) - d(n-1) - \mu K^2 e(n-1)$

$\Rightarrow H(z) = \dfrac{D(z)}{E(z)} = \dfrac{1 - z^{-1}}{1 - (1-\mu K^2) z^{-1}}$ $\Rightarrow$ stable for $|1 - \mu K^2| < 1$

$\Rightarrow 0 < \mu < \dfrac{2}{K^2}$

b)

$J'(n) = \frac{1}{2} e^2(n) + \frac{\beta}{2} \underline{w}^T(n) \underline{w}(n)$

↑ LMS cost

$\nabla\left(\frac{\beta}{2} \underline{w}^T \underline{w}\right)$

$\Rightarrow \nabla |e(n)|^2 = -e(n) \underline{x}(n) + \underline{w}(n) \cdot \frac{\beta}{2}$

$\Rightarrow \underline{w}(n+1) = w(n) + \mu e(n) \underline{x}(n) - \mu\beta \underline{w}(n) = [1 - \mu\beta] \underline{w}(n) + \mu e(n) \underline{x}(n)$

This cost function comprises the gradient update for the LMS algorithm $\Delta w_{LMS}(n) = \mu e(n) \underline{x}(n)$, and a penalty term $\frac{\beta}{2} \underline{w}^T \underline{w}$, which is quadratic in $\underline{w}$, which penalises for large $\underline{w}$.

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

$\Rightarrow \beta\mu \underline{w}(\infty) + \mu \underline{R}_x \underline{w}(\infty) = \mu \underline{r}_{dx} \Rightarrow (\underline{R}_x + \beta \underline{I}) \underline{w}(\infty) = \underline{r}_{dx}$

Therefore $\underline{w}(\infty) = [\underline{R}_x + \beta \underline{I}]^{-1} \underline{r}_{dx}$