EE4-13
EE9SO15

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2012

MSc and EEE/ISE PART IV: MEng and ACGI

SPECTRAL ESTIMATION AND ADAPTIVE SIGNAL PROCESSING

Thursday, 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) Consider power spectrum estimation of peaky spectra, for signals of length N.

a) Explain how periodogram based methods would perform in the estimation of peaky spectra, especially if there are two close spectral peaks of different magnitudes. [7]

b) Determine the segment length L in Bartlett's method of periodogram averaging, required to resolve the two consecutive spectral peaks which are Δf apart in frequency. For this value of L, find an aproximate value for the bias of the estimate at the peaks of the spectrum. Is this bias related to the area under the spectral peaks? [6]

Hint: Use the assumption that the width of the main lobe of the triangular (Bartlett) window $W_B(e^{j\omega})$ is much wider than the width of the spectral peaks, which allows us to assume $W_B(e^{j\omega}) \approx L$ over the interval $-\Delta\omega/2 \leq \omega \leq \Delta\omega/2$. The resolution of the standard periodogram is $\Delta\omega = 0.89\frac{2\pi}{N}$, where N is the number of data points.

c) The bandwidth and time-bandwidth product are key design parameters in spectrum estimation.

i) Let $W(\omega)$ denote a general spectral window (Fourier transform of the time-domain window $w(n)$) that has a peak at $\omega = 0$ and is symmetric about that point, and assume that the peak of $W(\omega)$ is narrow. Use a Taylor series expansion to show that an approximate formula for calculating the bandwidth B of the peak of $W(\omega)$ is [4]

$$B \approx 2\sqrt{|W(0)/W''(0)|}$$

Hint: The spectral bandwidth $B = \omega_2 - \omega_1$ is defined via the "half power points" for which $W(\omega_1) = W(\omega_2) = W(0)/2$, where $\omega_1 < \omega_2$.

Hint: Assume a second order Taylor series expansion around $\omega = 0$

$$W(\omega) \approx W(0) + W'(0)\omega + \frac{1}{2}W''(0)\omega^2$$

and take into account that at the peak $W'(0) = 0$.

ii) Use the formulas in Part i) to show that for a finite duration data window $w(n)$, which is nonzero only for $|n| < N$, the spectral peak bandwidth B satisfies [3]

$$B \times N \geq \frac{1}{\pi}$$

Hint: Express $|W''(0)|$ in terms of B and $W(0)$. For e.g. the Bartlett window, $W(0) = \sum_{n=-(N-1)}^{N-1} \frac{N-|n|}{N} e^{-j\omega n} = \frac{1}{N}\left|\sum_{n=0}^{N-1} e^{-j\omega n}\right|^2_{\omega=0} = N.$

2) Consider the filterbank interpretation of power spectrum estimation.

a) Sketch a block diagram of the filterbank interpretation of the periodogram. Explain the physical meaning of the transfer function of the sub-bands within such a filterbank (for each subfilter, $h_i(n) = \frac{1}{N}e^{jn\omega_i}w_R(n)$). [6]

b) Now consider the interpretation of the minimum variance power spectrum estimation method as a filterbank.

i) State the differences between the filterbank interpretations of the periodogram and the minimum variance method. Which one is data adaptive? [4]

ii) Explain in your own words the minimum variance method as a constrained optimisation problem and illuminate the physical meaning of the constraint involved, and that of the transfer functions of the sub-band filters. [4]

c) Now suppose that the sub-bands within the filterbank are modelled by ARMA(p,q) models. A minimum phase ARMA(p,q) model, described by its transfer function $H(z) = \frac{B(z)}{A(z)}$ can be equivalently represented as an $AR(\infty)$ model whose transfer function is given by $H(z) = \frac{1}{C(z)}$. Define

$$\begin{aligned} A(z) &= 1 + a_1 z^{-1} + \cdots + a_p z^{-p} \\ B(z) &= 1 + b_1 z^{-1} + \cdots + b_q z^{-q} \\ C(z) &= 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots \end{aligned}$$

i) Show that [3]

$$c_k = \begin{cases} 1, & k = 0 \\ a_k - \sum_{i=1}^{q} b_i c_{k-i}, & 1 \le k \le p \\ -\sum_{i=1}^{q} b_i c_{k-i}, & k > p \end{cases}$$

ii) Using the equations above, explain how you would compute the a_i and b_j parameters from a given set of $\{c_k\}_{k=0}^{p+q}$ parameters. Assume that p and q are known (some derivation needed). [3]

3) Consider autoregressive moving average (ARMA) power spectrum estimation.

a) For the ARMA(1,1) process $y(n)$ given by

$$y(n) = ay(n-1) + w(n) + bw(n-1)$$

show that the state-space representation for this process may be written as

$$\begin{aligned} \mathbf{x}(n) &= \begin{bmatrix} a & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x}(n-1) + \begin{bmatrix} 1 \\ b \end{bmatrix} w(n) \\ y(n) &= [1,\ 0]\, \mathbf{x}(n) \end{aligned}$$

where $\mathbf{x}(n) = [x_1(n), x_2(n)]^T$ is a 2-dimensional state vector. [6]

b) We would like to perform spectrum estimation of general ARMA processes.

i) Explain the benefits of ARMA power spectrum estimation over the AR and MA spectrum estimation. [3]

ii) Write down the equation for an ARMA power spectrum estimate. [3]

iii) Consider an ARMA power spectrum estimator $\hat{P}(\omega)$, where the ARMA parameters are real. Show that the asymptotic variance of this spectral estimator can be written in the form

$$E\Big\{ \big[\hat{P}(\omega) - P(\omega)\big]^2 \Big\} = C(\omega) P^2(\omega)$$

where $P(\omega)$ is the true power spectral density, and $C(\omega) = \mathbf{q}^T(\omega)\mathbf{P}\mathbf{q}(\omega)$. Here, $\mathbf{P} = E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\}$ is the covariance matrix of the estimate of the parameter vector $\boldsymbol{\theta} = [\sigma^2, \mathbf{a}, \mathbf{b}]$, where σ^2, $\mathbf{a}$ and $\mathbf{b}$ are respectively the driving noise variance and coefficient vectors of the AR and MA part of the ARMA model, and the vector $\mathbf{q}(\omega)$ needs to be found. [4]

Hint: For sufficiently long data, we can make use of a first order Taylor series expansion to write

$$\hat{P}(\omega) \approx P(\omega) + \frac{\partial P(\omega)}{\partial[\sigma^2, \mathbf{a}, \mathbf{b}]}[\hat{\sigma}^2 - \sigma^2, \hat{\mathbf{a}} - \mathbf{a}, \hat{\mathbf{b}} - \mathbf{b}]^T$$

c) Explain in your own words how you would design a time-frequency spectral estimator based on the state-space representation from Part a). What would be the trade-off in using such an estimator with respect to the data length considered. [4]

4) Consider a system consisting of two sensors, each making a single measurement of an unknown constant x. Each measurement is noisy and can be modelled as

$$y_1 = x + v_1 \qquad\qquad y_2 = x + v_2$$

where v_1 and v_2 are zero mean uncorrelated random variables with the respective variances σ_1^2 and σ_2^2 and y_1 and y_2 are the measured signals at the sensors.

a) In the absence of any other information, we seek the best linear estimate of x in the form

$$\hat{x} = k_1 y_1 + k_2 y_2$$

i) Find the values for k_1 and k_2 that yield an unbiased estimate of x that minimises the mean square error $E\{|x - \hat{x}|^2\}$ (Hint: express k_1 and k_2 in terms of σ_1^2 and σ_2^2). [4]

ii) Repeat Part i) for the case where the measurement errors are correlated, that is [4]

$$E\{v_1 v_2\} = \rho\sigma_1\sigma_2$$

wher ρ is the correlation coefficient.

b) If the system is to operate in real time, then combining the measurements and providing adaptive weighting to share cross-information would lead to a dual-channel least mean square (LMS) adaptive estimator given by

$$\begin{aligned}\hat{y}_1(n) &= a(n)y_1(n) + b(n)y_2(n)\\ \hat{y}_2(n) &= c(n)y_1(n) + d(n)y_2(n)\end{aligned}$$

where $a(n), b(n), c(n), d(n)$ are filter coefficients.

i) Based on the output errors $e_1(n) = x(n) - \hat{y}_1(n)$, $e_2(n) = x(n) - \hat{y}_2(n)$, and using the cost function

$$J(n) = \frac{1}{2}[e_1^2(n) + e_2^2(n)]$$

derive the weight updates for this multichannel LMS algorithm. [4]

ii) Explain how this estimator operates when estimating only a single variable x as compared to the usual way of estimating two different variables, say, x_1 and x_2. [4]

iii) Use a complex valued filter for the same problem, where $z(n) = y_1(n) + jy_2(n)$. Comment on the circularity properties and the expected performance, as compared with Part i). Would the strictly linear CLMS or widely linear ACLMS be an optimal solution, and explain what opportunities the complex approach offers? [4]

5) Adaptive filters deal with noise in several different ways.

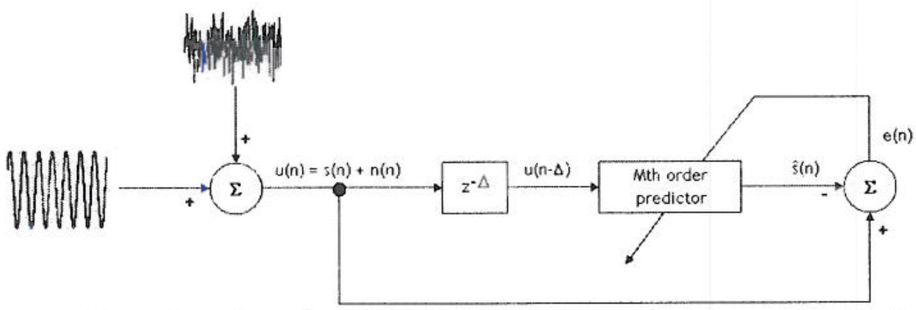
a) Explain how an adaptive filter is connected to the environment and draw a block diagram of the noise cancelling configuration. Explain the operation of a noise canceller and identify some applications where this scheme would be useful. [6]

b) Suppose the input to an adaptive predictor is white noise with an auto-correlation sequence $r_x(k) = \sigma_x^2 \delta(k)$.

i) Find the Wiener solution for a one-step ahead predictor of white noise. [4]

ii) Derive the method of steepest descent and minimise the mean square prediction error using the method of steepest descent with a step size $\mu = 1/(5\sigma_x^2)$ and an initial weight vector $\mathbf{w}(0) = [1, 1, 1, 1]^T$. Does the method of steepest descent converge to the solution found in Part i)? [6]

c) When filtering real world data corrupted by white noise, explain how the following "line enhancer" configuration would remove noise. Does this



configuration allow the additive noise to be correlated, and how is the noise correlation width related to the delay Δ? Explain the role of the teaching signal. [4]

Spectral Estimation + Adaptive Signal Processing EE4-13, EE9SO15

①

1/8

SOLUTIONS: 2012

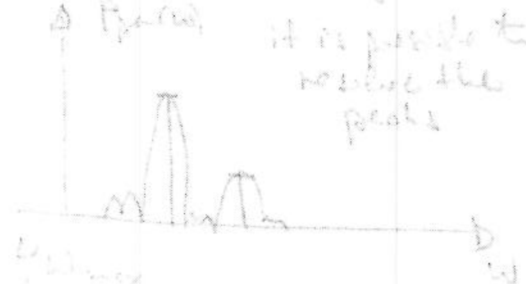
1.

a) The periodogram is given by

$$\hat{P}_{per}(f) = \frac{1}{N}\left|\sum_{k=0}^{N-1} x(k)e^{-j2\pi fk}\right|^2$$

It is an MA method and Fourier based. Being an MA method, there are no poles in the transfer function and peaky spectra are not well estimated from short data segments. For $N \to \infty$, the periodogram is a good estimator of peaky spectra ([illegible] from [illegible] MA model). If two [illegible] [illegible] [illegible].

for a short N, they are likely to be meshed. For a large N



it is possible to resolve the peaks

The condition is

$$\Delta\omega > 2\frac{2\pi}{N} \quad \ldots \quad N \geq \frac{4\pi}{\Delta\omega}$$

b) For resolving two peaks that are separated by Δf, we require the segment length $L \geq \frac{0.89}{\Delta f}$

The bias $E\{\hat{P}_B(e^{j\omega})\} = E\left[\frac{1}{K}\sum_{k=1}^{K}\hat{P}_{per}^{(k)}(e^{j\omega})\right] = E\left[\hat{P}_{per}^{(k)}(e^{j\omega})\right]$

and thus $E\{\hat{P}_B(e^{j\omega})\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\theta})W_B(e^{j(\omega-\theta)})d\theta$

where $W_B = \frac{1}{L}\left[\frac{\sin(L\omega/2)}{\sin(\omega/2)}\right]^2$. Since $W_B(e^{j\omega})$ has been selected so that the two peaks can be resolved, let's assume that $W_B(e^{j\omega})$ is nonzero over $-\frac{\Delta\omega}{2} \leq \omega \leq \frac{\Delta\omega}{2}$. Furthermore, since the width of the [illegible] [illegible] is much wider than the width of spectral peaks, if we assume that $W_B(e^{j\omega}) \simeq 1$ over the interval $-\frac{\Delta\omega}{2} \leq \omega \leq \frac{\Delta\omega}{2}$, then we can form the following approximation:

$$E\{\hat{P}_B(e^{j\omega})\} \simeq \frac{1}{2\pi} \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} P_x(e^{j(\omega-\theta)})\, d\theta$$

Therefore, $E\{\hat{P}_B(e^{j\omega})\}$ ~~[illegible]~~ at $\omega = \omega_1$ and $\omega = \omega_2$ is proportional to the area under the spectral peaks.

c) As $\omega_2 \leq \pi$ (as the spectral peak is narrow), we can use
i) the Taylor series expansion around $\omega \simeq 0$ to write:

$$W(\omega) \simeq W(0) + W'(0)\omega + \frac{1}{2} W''(0)\omega^2 = W(0) + \frac{1}{2} W''(0)\omega^2$$

where the first [illegible]

$W'(0) = 0$

For the Bartlett Window,

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

~~[illegible]~~

and
$$W''(0) = \frac{2}{N} \mathrm{Re}\left[\left(-\sum_{n=0}^{N-1} n^2 e^{j\omega n} \sum_{n=0}^{N-1} e^{-j\omega n}\right) + \left|\sum_{n=0}^{N-1} n e^{j\omega n}\right|^2\right]_{\omega=0}$$

$$= \frac{2}{N}\left(-\frac{N(N-1)(2N-1)}{6} N + \frac{N^2(N-1)^2}{4}\right) = -\frac{N(N-1)(N+1)}{6} \simeq -\frac{N^3}{6}$$

which leads to the following bandwidth (rad/sample)

[illegible]

Hence $\frac{\sqrt{6}}{\pi N} \simeq 0.78/N$

This bandwidth is underestimated, as we truncated the TS, however the neglected terms contribute to the error when $N \to \infty$.

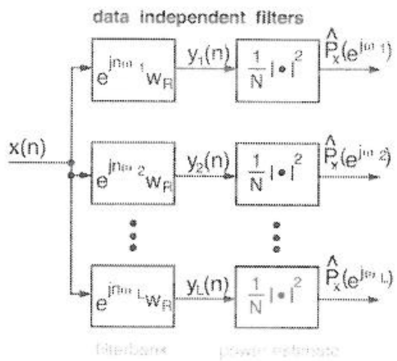
ii) A straightforward calculation shows that:

$$|W''(0)| = \left|-\sum_{k=-(N-1)}^{N-1} k^2 w(k)\right| \leq \left|\sum_{k=-(N-1)}^{N-1} N^2 w(k)\right| \leq N^2 W(0)$$

[w, not W] [w, not W]

Hence $BN \geq \frac{1}{\pi}$.

2) [**bookwork and new examples**]
The filterbank interpretation of the periodogram is shown in the figure below.



Since

$$h_i(n) = \frac{1}{N} e^{\jmath n\omega_i} w_R(n) = \begin{cases} \frac{1}{N} e^{\jmath n\omega_i} & ; \quad 0 \le n \le N \\ 0 & ; \quad \text{otherwise} \end{cases}$$

The frequency response is

$$H_i(e^{\jmath\omega}) = \sum_{n=0}^{N-1} h_i(n) e^{-\jmath n\omega} = e^{-\jmath(\omega-\omega_i)\frac{(N-1)}{2}} \frac{\sin[N(\omega-\omega_i)/2]}{N \sin[(\omega-\omega_i)/2]}$$

If we force $|H_i(e^{\jmath\omega_i})| = 1 \;\Rightarrow\; P_x(e^{\jmath\omega_i}) = P_y(e^{\jmath\omega_i})$, then $P_x(e^{\jmath\omega_i}) \approx NE\{|y_i(n)|^2\}$, explaining the principle. The sub-band transfer functions are data-independent sync functions, which provide a physical interperation of the periodogram performance.

b) i) and ii)

- Periodogram is formed by dividing this power estimate by the filter bandwidth $\Delta = 2\pi/N$, and
$$\hat{P}_x(e^{\jmath\omega_i}) = \frac{E\left\{|y_i(n)|^2\right\}}{\Delta/2\pi}$$

- Each filter in the filter bank of a periodogram is the same (errrr apart from the centre frequency) $\Rightarrow$ these filters are **data independent**

- Result $\Rightarrow$ When a random process contains a significant amount of power in frequency bands within the sidelopbes of the bandpass filter, leakage through the sidelobes will lead to significant distortion in the power estimates

- **Solution:** allow each filter in the filter bank to be **data adaptive** and **optimum** in the sense of rejecting as much out–of–band signal power as possible

The minimum variance (MV) spectum estimation is based on this idea and involves the following steps:

1. Design a bank of bandpass filters $g_i(n)$ with venter frequency ω_i so that each filter rejects the maximum amount of out–of–band power while passing the component at frequency ω_i with no distortion

2. Filter $x(n)$ with each filter in the filter bank and estimate the power in each output process $y_i(n)$

3. Set

$$\hat{P}_x(e^{\jmath\omega_i}) = \frac{E\left\{|y_i(n)|^2\right\}}{\Delta/2\pi}$$

that is power estimated from step (2) divided by the filter bandwidth.

To achieve this the filterbank transfer functions need to be as close to the ideal bandpass filter as possible, $G_i(\omega)$ will be constrained to have gain one at $\omega = \omega_i$, that is $G_i(\omega) = \sum_{n=0}^{p} g_i(n)e^{-\jmath n\omega_i} = 1$. For $\mathbf{g}_i = [g_i(0), \ldots, g_i(p)]^T$ vector of filter coefficients and $\mathbf{e}_i = [1, e^{\jmath\omega_i}, \ldots, e^{\jmath p\omega_i}]^T$ vector oo complex exponentials, the above constraint results in $\mathbf{g}_i^H \mathbf{e}_i = 1$.

c) i) Since $B(z)/A(z) = 1/C(z)$ we have $A(z) = B(z)C(z)$, so $a_k = \sum_{i=0}^{p} b_i c_{k-i}$. Thus we immediately have the answer.

ii) Writing $c_k = -\sum_{i=1}^{p} b_i c_{k-i}$ for $q+1 \leq k \leq p+q$ in a matrix form, allows to solve for b_i, assuming that the matrix inverse exists. Then the a_i coefficients are obtained as

$$a_k = c_k + \sum_{i=1}^{q} b_i c_{k-i}$$

(3) (a) $\underline{x}(n) = \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} \Rightarrow \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \begin{bmatrix} a & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} + \begin{bmatrix} 1 \\ b \end{bmatrix} w(n)$

Thus, $x_1(n) = a x_1(n-1) + x_2(n-1) + w(n)$
$x_2(n) = b w(n)$

and $x_1(n) = a x_1(n-1) + b w(n-1) + w(n)$, which with $y(n) = [1 \;\; 0]\, x(n)$ is the equation of the ARMA(1) process.

(b) The ARMA spectrum models both peaky spectra and spectra with zeros.

(ii) From $H(z) = \dfrac{A(z)}{B(z)} = \dfrac{b_0 + b_1 z^{-1} + \dots + b_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}}$

we have $P_{ARMA}(e^{j\omega}) = \dfrac{\left|\sum_{k=0}^{q} b_q(k) e^{-jk\omega}\right|^2}{\left|1 + \sum_{k=1}^{p} a_p(k) e^{-jk\omega}\right|^2}$

(iii) From

$$\hat{P}(\omega) \approx P(\omega) + \frac{\partial P(\omega)}{\partial [\sigma^2, \underline{a}, \underline{b}]} \begin{bmatrix} \hat{\sigma}^2 - \sigma^2 \\ \hat{\underline{a}} - \underline{a} \\ \hat{\underline{b}} - \underline{b} \end{bmatrix}$$

Introduce the notation

$$P = E\left\{ \begin{bmatrix} \hat{\sigma}^2 - \sigma^2 \\ \hat{\underline{a}} - \underline{a} \\ \hat{\underline{b}} - \underline{b} \end{bmatrix} \left[\hat{\sigma}^2 - \sigma^2, (\hat{\underline{a}} - \underline{a})^T, (\hat{\underline{b}} - \underline{b})^T \right] \right\}$$

It then follows that for large data sets

$$E\{\hat{P}(\omega) - P(\omega)\}^2 \approx [\underline{g}^T(\omega) P \underline{g}(\omega)] P^2 = C(\omega) P^2(\omega)$$

where $\underline{g}$ depends on the variance of the driving noise and coefficients $\underline{a}$ and $\underline{b}$.

c) A time-frequency spectrogram may be designed by calculating the PSD sequentially, based on the ARMA spectrum in Part a) or otherwise, and plotting it along the time

Short N = poor resolution
Long N = large computational complexity

4 (a i) For an unbiased estimate, we want

$$E\{x-\hat{x}\} = E\{x\} - k_1 E\{y_1\} - k_2 E\{y_2\} = 0$$

Since v_1 and v_2 are zero mean, and $E\{x\} = x$ ← constant

$$\Rightarrow E\{x-\hat{x}\} = x - (k_1+k_2)x = 0$$

Therefore, we want:

$$k_1 + k_2 = 1 \quad \text{or} \quad k_2 = 1-k_1$$

The MSE that we want to minimise is

$$E\{[x-\hat{x}]^2\} = E\{[x - k_1 y_1 - (1-k_1) y_2]^2\} =$$

$$= E\{[-k_1 v_1 - (1-k_1) v_2]^2\} = k_1 \sigma_1^2 + (1-k_1)^2 \sigma_2^2$$

(since v_1 and v_2 are uncorrelated)

$\Rightarrow$ To find k_1 that minimizes the MSE

$$\frac{\partial}{\partial k_1} E\{[x-\hat{x}]^2\} = 2k_1\sigma_1^2 - 2(1-k_1)\sigma_2^2 = 0$$

Solve to find $k_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ and $k_2 = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$

$$\Rightarrow \boxed{\hat{x} = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} y_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} y_2}$$

ii) The only change comes when we evaluate the MSE, giving

$$E\{[x-\hat{x}]^2\} = E\{[-k_1 v_1 - (1-k_1) v_2]^2\} = k_1\sigma_1^2 + 2k_1(1-k_1)\rho\sigma_1\sigma_2 + (1-k_1)^2\sigma_2^2$$

SET $\frac{\partial}{\partial k_1} E\{[x-\hat{x}]^2\} = 2k_1\sigma_1^2 + 2(1-2k_1)\rho\sigma_1\sigma_2 - 2(1-k_1)\sigma_2^2 = 0$

$$\Rightarrow k_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \qquad k_2 = 1-k_1 = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$$

b) The dual channel LMS

(i) Based on the LMS, we have

$$a(n+1) = a(n) + \mu e_1(n) y_1(n)$$
$$b(n+1) = b(n) + \mu e_1(n) y_2(n)$$
$$c(n+1) = c(n) + \mu e_2(n) y_1(n)$$
$$d(n+1) = d(n) + \mu e_2(n) y_2(n)$$

(Since the filter length $L=1$)

This would be [illegible]

based on

$$\Delta \underline{w} = -\mu \nabla_{\underline{w}} J,$$

iii) $\underline{w} \in (\underline{a}, \underline{b}, \underline{c}, \underline{d})$

(ii) This reflects the principle of distributed, or cooperative estimation of a single parameter by many sensors, and a basic update of a sensor network algorithm.

This would allow the error variance to be halved, as e.g. comparing with the estimation of a single channel, or two different quantities, x_1, x_2.

iii) A complex representation would give:

CLMS: $\underline{w}(n+1) = \underline{w}(n) + \mu e x^*$

ACLMS: $\Delta \underline{w} = \mu e x^* + \mu e^* x$

$$z(n) = y_1(n) + j y_2(n) = x + jx + v_1 + j v_2$$

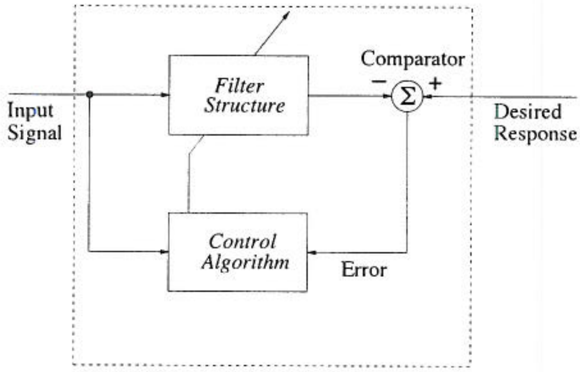
Since the powers of x are the same $\Rightarrow x$ is circular

In general $\sigma_{v_1}^2 \neq \sigma_{v_2}^2 \Rightarrow$ noncircular, and ACLMS is optimal.

OPPORTUNITY! Physical meaning of circularity coeff.

5) [**bookwork, new example, independent reasoning**]

a) Generic adaptive filter



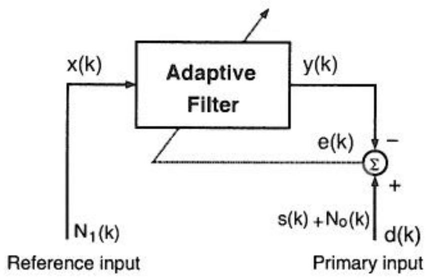
The filter is connected via its input, output and teaching signal, and has an internal variable (the error) which is used in the filter adaptation. The main design paramteres are:

Filter architecture: (FIR, IIR, linear, nonlinear)

Filter function: prediction, system identification, inverse system modelling, noise cancellation

Adaptation: based on e^2, $|e|$, e^4, etc

The adaptive noise canceller in its standard form removes the noise from the useful input, by comparing it statistically with the noise in the 'reference input'. The only requirement for the reference noise is to be correlated with the noise corrupting the signal. The higher the correlation the better the performance.



A typical application would be in biomedical engineering (removal or reaspiratory sounds from heart electrical activity, separation of maternal and foetal ECG) and in data communications, e.g. over the twisted wire pair.

b) i) With

$$\mathbf{R}_x = \sigma_x^2 \mathbf{I} \qquad \text{and} \qquad \mathbf{r}_{dx} = \mathbf{0}$$

the solution to the Wiener-Hopf equations is $\mathbf{w} = 0$.

ii) The steepest descent algorithm in its vector-matrix form is (show derivation)

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

8/8

Since $\mathbf{R}_x = \sigma_x^2\mathbf{I}$ and $\mathbf{r}_{dx} = \mathbf{0}$, then

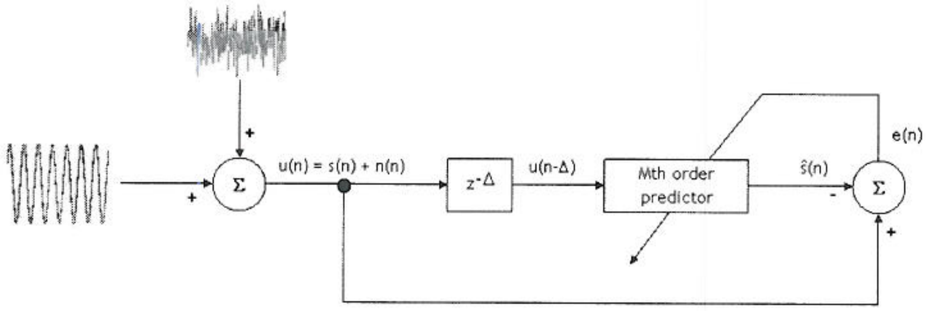
$$\mathbf{w}(n+1) = (1 - \mu\sigma_x^2)\mathbf{w}(n)$$

With $\mu = 1/(5\sigma_x^2)$, the time evolution of $\mathbf{w}(n)$ becomes

$$\mathbf{w}(n) = (1 - \frac{1}{5})^n\mathbf{w}(0)$$

which goes to zero as $n \to \infty$, that is, it gives asymptotically the same solution as the Wiener filter in Part i).

c) The noise cancelling configuration in the figure below



is suitable when the noise is white or has very narrow correlation structure. In that case, the *signal + noise* which serves as a "teaching signal" contains both signal and noise, whereas the "input" is the noisy signal delayed by Δ samples. If Δ is larger than the correlation structure in the noise, then it is only the useful signal that is correlated and the scheme operates similarly to the noise cancellation scheme in a). For white noise, ideally $\Delta = 1$. If the noise is correlated, with narrower correlation structure than the signal, then we can still use this scheme, as long as the correlation structure of the noise is smaller than Δ.