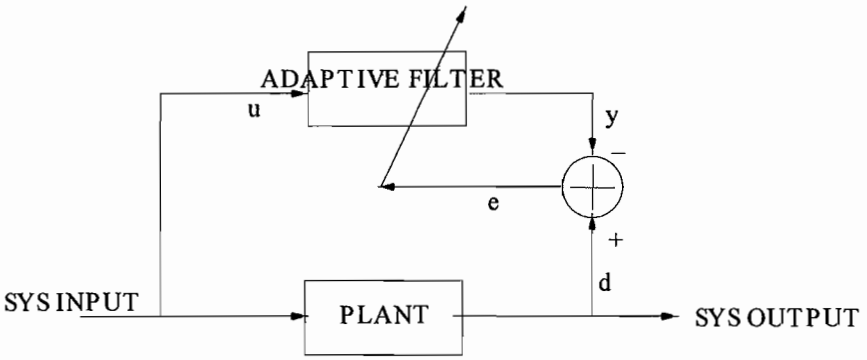
ENGINEERING TRIPOS PART IIB
Module 4F7: Digital Filters and Spectrum Estimation
Monday 20 April 2009, 9 to 10.30
Solutions

May 11, 2009

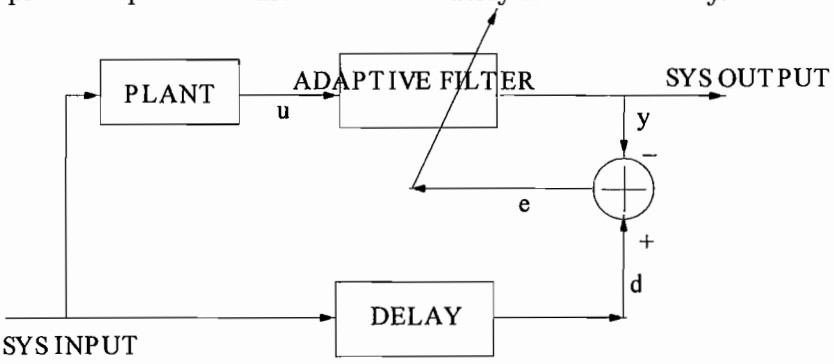
1 Question One

Part (a)

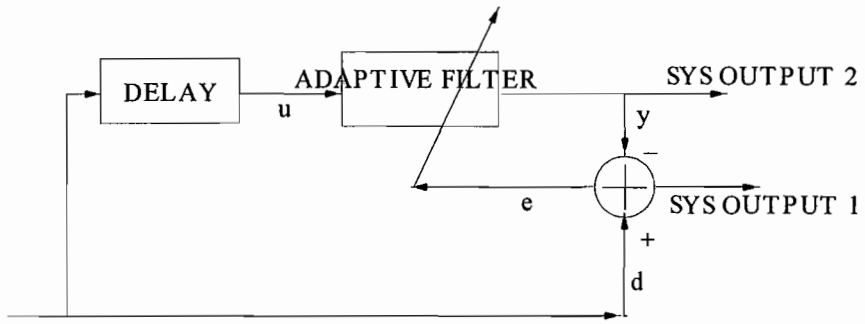
Identification: A mathematical model is an essential component for analysis and control. An adaptive filter is used to provide a linear model that represents the best fit (in some sense) to an unknown *plant*. The plant and adaptive filter are driven by the same input and the plant output supplies the desired response. If the plant is dynamic in nature, the adaptive filter may be able to track the time-varying plant model provided it changes slower than the adaptation rate.



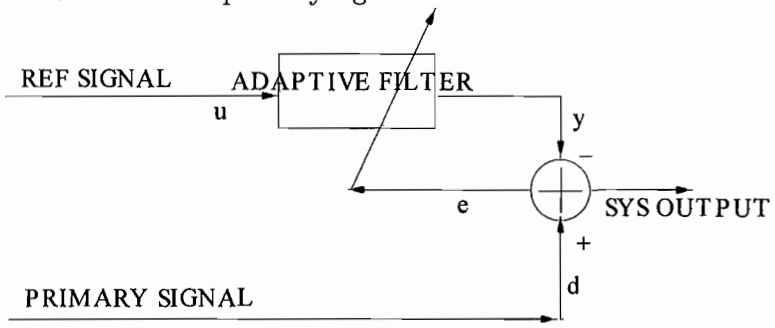
Inverse modeling: The function of the adaptive filter here is to provide the best fitting (in some sense) *inverse* model to the *unknown noisy plant.* Ideally the inverse model will have a transfer function equal to the inverse of the plant's transfer function. The desired response is a delayed version of the plant's input. In some instances a delay is not necessary.



Prediction: The function of the adaptive filter is to provide the best prediction (in some sense) of the present value of the input signal. The input of the filter are the past values, hence the delay. The desired response is the present value.



Interference cancellation: The adaptive filter is used to cancel interference which is present alongside the signal of interest in the *primary* signal. The input to the filter is the *reference* signal. This signal is derived from a sensor located relative to the sensor supplying the primary signal in such a way that the signal of interest is undetectable. The adaptive filter will synthesize the interference in the primary signal using the reference signal which is then subtracted from the primary signal.



Part (b)

$$\mathbf{R} = E\left\{[u(n) \quad u(n-1)]^{\mathrm{T}}[u(n) \quad u(n-1)]\right\}$$

$$= E\left\{\begin{bmatrix} u(n)u(n) & u(n)u(n-1) \\ u(n-1)u(n) & u(n-1)u(n-1) \end{bmatrix}\right\}$$

$$= \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}$$

Solve for eigenvalues: $det(\begin{bmatrix} 1-\lambda & \alpha \\ \alpha & 1-\lambda \end{bmatrix}) = 0$. $(1-\lambda)^2 - \alpha^2 = 0$. This gives

$$1 - \lambda = \pm\alpha \text{ or } \lambda = 1 \pm \alpha$$

Part (c)

$$\begin{aligned} S_u(e^{j\omega}) &= \sum_{k=-\infty}^{k=\infty} r_u(k)e^{-j\omega k} \\ &= \sum_{k=-\infty}^{0} \alpha^{|k|}e^{-j\omega k} + \sum_{k=0}^{\infty}\alpha^{|k|}e^{-j\omega k} - r_u(0) \\ &= \frac{1}{1-\alpha e^{j\omega}} + \frac{1}{1-\alpha e^{-j\omega}} - 1 \end{aligned}$$

Now differentiate this function with respect to ω and set the derivative to 0:

$$\frac{d}{d\omega}S_u(e^{j\omega}) = \frac{-1}{(1-\alpha e^{j\omega})^2}(-\alpha j e^{j\omega}) + (-1)\frac{1}{(1-\alpha e^{-j\omega})^2}(\alpha j e^{-j\omega})$$

Setting the derivative to zero:

$$\frac{e^{j\omega}}{(1-\alpha e^{j\omega})^2} = \frac{e^{-j\omega}}{(1-\alpha e^{-j\omega})^2}$$

$\omega = 0, \pi$ solves this equation. So

$$\begin{aligned} S_u(e^{j0}) &= \frac{1}{1-\alpha} + \frac{1}{1-\alpha} - 1 = \frac{1+\alpha}{1-\alpha} \\ S_u(e^{j\pi}) &= \frac{1}{1+\alpha} + \frac{1}{1+\alpha} - 1 = \frac{1-\alpha}{1+\alpha} \end{aligned}$$

If $\alpha > 0$, $\lambda_{\max} = \frac{1+\alpha}{1-\alpha}$, $\lambda_{\min} = \frac{1-\alpha}{1+\alpha}$. If $\alpha < 0$, $\lambda_{\min} = \frac{1+\alpha}{1-\alpha}$, $\lambda_{\max} = \frac{1-\alpha}{1+\alpha}$.

2 Question Two

Part (a)

The M-tap LMS algorithm for linear prediction is

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu(u(n) - \mathbf{h}(n)^{\mathrm{T}}\mathbf{u}(n-1))\mathbf{u}(n-1)$$

where $\mathbf{u}(n-1) = [u(n-1), u(n-2), \ldots, u(n-M)]$.

Part (b) Step-size range is $0 < \mu < \frac{2}{\lambda_{\max}}$ where $\lambda_{\max}$ is the maximum eigenvalue of matrix $\mathbf{R}$ defined as follow:

$$\begin{aligned}\mathbf{R} &= E\{\mathbf{u}(n)\mathbf{u}(n)^{\mathrm{T}}\} \\ &= E\left\{(\mathbf{x}(n)+\mathbf{v}(n))(\mathbf{x}(n)+\mathbf{v}(n))^{\mathrm{T}}\right\} \\ &= E\left\{\mathbf{x}(n)\mathbf{x}(n)^{\mathrm{T}}+\mathbf{x}(n)\mathbf{v}(n)^{\mathrm{T}}+\mathbf{v}(n)\mathbf{x}(n)^{\mathrm{T}}+\mathbf{v}(n)\mathbf{v}(n)^{\mathrm{T}}\right\} \\ &= \mathbf{R}_x+\sigma_v^2\mathbf{I}\end{aligned}$$

where $\mathbf{I}$ is the $M \times M$ identity matrix.

We know the LMS will converge in mean to the same solution as the Steepest Descent algorithm which is

$$\lim_{n\to\infty} E\{\mathbf{h}(n)\} = (\mathbf{R}_x+\sigma_v^2)^{-1}E\{\mathbf{u}(n-1)u(n)\}$$

and

$$\begin{aligned}\mathbf{p} &= E\{\mathbf{u}(n-1)u(n)\} = E\{(\mathbf{x}(n-1)+\mathbf{v}(n-1))(x(n)+v(n))\} \\ &= E\{(\mathbf{x}(n-1)x(n)\}\end{aligned}$$

Part (c) The γ-LMS is

$$\mathbf{h}(n+1) = \gamma\mathbf{h}(n)+\mu(u(n)-\mathbf{h}(n)^{\mathrm{T}}\mathbf{u}(n-1))\mathbf{u}(n-1)$$

Take expectation on both sides:

$$\begin{aligned}E\{\mathbf{h}(n+1)\} &= \gamma E\{\mathbf{h}(n)\}+\mu E\left\{(u(n)\mathbf{u}(n-1)-\mathbf{u}(n-1)\mathbf{u}(n-1)^{\mathrm{T}}\mathbf{h}(n))\right\} \\ &= \gamma E\{\mathbf{h}(n)\}+\mu\mathbf{p}-\mu E\left\{\mathbf{u}(n-1)\mathbf{u}(n-1)^{\mathrm{T}}\mathbf{h}(n)\right\} \\ &\approx \gamma E\{\mathbf{h}(n)\}+\mu\mathbf{p}-\mu E\left\{\mathbf{u}(n-1)\mathbf{u}(n-1)^{\mathrm{T}}\right\}E\{\mathbf{h}(n)\} \\ &= \gamma E\{\mathbf{h}(n)\}+\mu\mathbf{p}-\mu(\mathbf{R}_{\mathbf{x}}+\sigma_v^2\mathbf{I})E\{\mathbf{h}(n)\}\end{aligned}$$

The Independence Assumption was invoked in the second last line. Call the limit point $\mathbf{h}$:

$$\mathbf{h} = \gamma\mathbf{h}+\mu\mathbf{p}-\mu(\mathbf{R}_{\mathbf{x}}+\sigma_v^2\mathbf{I})\mathbf{h}$$

We would like the limit point to satisfy $\mathbf{R}_{\mathbf{x}}\mathbf{h} = \mathbf{p}$. This would require

$$\mathbf{h} = \gamma\mathbf{h}-\mu\sigma_v^2\mathbf{h}$$

or

$$1 = \gamma-\mu\sigma_v^2.$$

Thus, assuming the algorithm converges in the mean, choosing $\mu = (\gamma-1)/\sigma_v^2$ and $\gamma > 1$ will remove the bias. (We have not specified further restriction on γ to ensure convergence in mean.)

3 Question Three

Part (a): The random process $\{X_n\}$ is WSS if the following 3 conditions are satisfied:

1) Constant mean

2) The autocorrelation function $E(X_n X_m)$ depends on only the difference $|m-n|$

3) The variance of the process, $E(X_n^2) - E(X_n)^2$, is finite.

The Power Spectrum is the Discrete Time Fourier Transform (DTFT) of the autocorrelation function $R_{XX}[k] = E(X_n X_{n+k})$:

$$S_X(e^{j\omega}) = \sum_{k=-\infty}^{\infty} R_{XX}[k](e^{-j\omega})^k$$

Part (b): Let the data points be $\{x_0, x_1, \ldots, x_{N-1}\}$.

The Correlogram estimate for the power spectrum is obtained by taking the DTFT of the sample autocorrelation function:

$$\hat{S}_X(e^{j\omega}) = \sum_{k=-L}^{L} \hat{R}_{XX}[k](e^{-j\omega})^k$$

where $L << N$ and $\hat{R}_{XX}[k]$ is either the biased or unbiased estimate:

$$\hat{R}_{XX}[k] = \frac{1}{N-k} \sum_{n=0}^{N-1-k} x_n x_{n+k}, \qquad 0 \leq k < N \qquad \text{(unbiased)}$$

$$\hat{R}_{XX}[k] = \frac{1}{N} \sum_{n=0}^{N-1-k} x_n x_{n+k}, \qquad 0 \leq k < N \qquad \text{(biased)}$$

In both cases we set $\hat{R}_{XX}[-k] = \hat{R}_{XX}[k]$.

The Periodogram estimates uses $L = N-1$ and the biased estimate for R_{XX}. In this case, the well known simplification is:

$$\hat{S}_X(e^{j\omega}) = \frac{1}{N}|X_{\text{win}}(e^{j\omega})|^2,$$

$$X_{\text{win}}(e^{j\omega}) = \sum_{n=0}^{N-1} x_n (e^{-j\omega})^n.$$

This shows the Periodogram estimator is non-negative.

Part (c):

$$\begin{aligned} E\{\hat{S}_X(e^{j\omega})\} &= \sum_{k=-N+1}^{N-1} E\{\hat{R}_{XX}[k]\}(e^{-j\omega})^k \\ &= \sum_{k=-N+1}^{N-1} \frac{N-|k|}{N} R_{XX}[k](e^{-j\omega})^k \\ &= DTFT(w_k \times R_{XX}[k]) \end{aligned}$$

where $w_k = \frac{N-|k|}{N}$, $-N+1 \leq k \leq N-1$, and zero for values of k beyond this range. Thus

$$E\{\hat{S}_X(e^{j\omega})\} = \frac{1}{2\pi} W(e^{j\omega}) * S_X(e^{j\omega})$$

where $W(e^{j\omega})$ =DTFT(w_k).

Part (d) For zero mean white noise, $E(X_n) = 0$, $E(X_n^2) = \sigma^2$, $E(X_n X_{n+k}) = 0$ when $k \neq 0$. Using the result from part (c),

$$\begin{aligned} &E\{\hat{S}_X(e^{j\omega})\} \\ &= \sum_{k=-N+1}^{N-1} \frac{N-|k|}{N} R_{XX}[k](e^{-j\omega})^k \\ &= R_{XX}[0] = \sigma^2. \end{aligned}$$

So the estimate is unbiased.

Part (e). The Blackman-Tukey method applies a window function of length $2L+1$ to the estimated autocorrelation function:

$$\begin{aligned} \hat{S}_X^{BT}(e^{j\omega}) &= \sum_{k=-L}^{L} v_k \hat{R}_{XX}[k](e^{-j\omega})^k \\ &= \frac{1}{2\pi} V(e^{j\omega}) * \hat{S}_X(e^{j\omega}) \end{aligned}$$

where $L < N-1$, $\hat{S}_X(e^{j\omega})$ is the Periodogram and v_k is a suitable window function. The expected value of the estimate is

$$E\{\hat{S}_X^{BT}(e^{j\omega})\} = \frac{1}{2\pi} V(e^{j\omega}) * E\{\hat{S}_X(e^{j\omega})\}$$

The variance is reduced by diminishing the contribution of $\hat{R}_{XX}[k]$ for values of k close to N to the power spectrum estimate. There are only very

few data samples that are used to estimate $\hat{R}_{XX}[k]$ for values of k close to N and these estimates can have high variance

The drawback is that the spectrum estimate is further smoothed by the applied window function and there is a loss in frequency resolution.

4 Question Four

Part (a) (A solution that includes all the points below would be ideal.)

If the physical process which generated the data is known or can be well approximated, then a parametric model can be constructed.
Careful estimation of the parameters in the model can lead to power spectrum estimates with **improved bias/variance** over non-parametric methods, e.g. periodogram-based methods.
We will consider spectrum estimation for LTI systems driven by a white noise input sequence. If a random process $\{X_n\}$ can be modeled as white noise exciting a filter with frequency response $H(e^{j\omega T})$ then the spectral density of the data can be expressed as:

$$S_X(e^{j\omega T}) = \sigma_w^2 \, |H(e^{j\omega T})|^2$$

where σ_w^2 is the variance of the white noise process.

We will study models in which the frequency response $H(e^{j\omega T})$ can be represented by a **finite number of parameters** which are estimated from the data.

Parametric models need to be chosen carefully - an inappropriate model for the data can give misleading results

A quite general representation is the autoregressive moving-average (ARMA) model. The ARMA(P,Q) model difference equation representation is:

$$x_n = -\sum_{p=1}^{P} a_p \, x_{n-p} + \sum_{q=0}^{Q} b_q \, w_{n-q} \tag{1}$$

where:

a_p are the AR parameters,
b_q are the MA parameters

and $\{W_n\}$ is a zero-mean white noise process with unit variance, $\sigma_w^2 = 1$. The ARMA model is a pole-zero IIR filter-based model with transfer function

$$H(z) = \frac{B(z)}{A(z)}$$

where:

$$A(z) = 1 + \sum_{p=1}^{P} a_p z^{-p}, \quad B(z) = \sum_{q=0}^{Q} b_q z^{-q}$$

We will always assume that the filter is stable, i.e. the poles (solutions of $A(z) = 0$) all lie **within** the unit circle

The power spectrum of the ARMA process is:

$$S_X(e^{j\omega T}) = \frac{|B(e^{j\omega T})|^2}{|A(e^{j\omega T})|^2}$$

The ARMA model is quite a flexible and general way to model a stationary random process: The poles model well the *peaks* in the spectrum (sharper peaks implies poles closer to the unit circle) The zeros model troughs in the spectrum. Complex spectra can be approximated well by large model orders P and Q

Part (b)

The AR model can be written equivalently as:

$$x_n = -\sum_{p=1}^{P} a_p \, x_{n-p} + e_n. \tag{2}$$

A convenient way to fit an AR model to the data is by choosing the parameter vector $\mathbf{a}$ which minimizes the total squared prediction error, E:

$$\mathcal{E} = \sum_{n=n_I}^{n_F} e_n^2 = \mathbf{e}^{\mathrm{T}}\mathbf{e}$$

where $\mathbf{e} = [e_{n_I}, e_{n_I+1}, , e_{n_F}]^{\mathrm{T}}$. In matrix notation: $\mathbf{e} = \mathbf{x} + \mathbf{X}\mathbf{a}$ where

$$\mathbf{x} = \begin{bmatrix} x_{n_I} \\ x_{n_I+1} \\ \vdots \\ x_{n_F} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_{n_I-1} & x_{n_I-2} & \dots & x_{n_I-P} \\ x_{n_I} & x_{n_I-1} & \dots & x_{n_I-P+1} \\ \vdots & \vdots & & \vdots \\ x_{n_F-1} & x_{n_F-2} & \dots & x_{n_F-P} \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix} \tag{3}$$

Now, expand $\mathbf{e}^{\mathrm{T}}\mathbf{e}$ and differentiate:

$$\mathbf{e} = \mathbf{x} + \mathbf{X}\,\mathbf{a}$$

$$\begin{aligned}\mathbf{e}^{\mathrm{T}}\mathbf{e} &= (\mathbf{x} + \mathbf{X}\,\mathbf{a})^{\mathrm{T}}(\mathbf{x} + \mathbf{X}\,\mathbf{a}) \\ &= \mathbf{x}^{\mathrm{T}}\mathbf{x} + 2\mathbf{x}^{\mathrm{T}}\mathbf{X}\mathbf{a} + \mathbf{a}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{a}\end{aligned}$$

$$\begin{aligned}\frac{\partial(\mathbf{e}^{\mathrm{T}}\mathbf{e})}{\partial\mathbf{a}} &= 2\frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{X}\mathbf{a})}{\partial\mathbf{a}} + \frac{\partial(\mathbf{a}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{a})}{\partial\mathbf{a}} \\ &= 2\mathbf{X}^{\mathrm{T}}\mathbf{x} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\,\mathbf{a}\end{aligned}$$

For the minimum of $\mathbf{e}^{\mathrm{T}}\mathbf{e}$:

$$\frac{\partial(\mathbf{e}^{\mathrm{T}}\mathbf{e})}{\partial\mathbf{a}} = \mathbf{0}$$

Therefore, $\mathbf{a} = -(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{x}$.

The covariance method minimizes only those error terms which can be fully calculated from the data. Examine the error equation:

$$e_n = x_n + \sum_{p=1}^{P} a_p\, x_{n-p}$$

The first error term that can be fully calculated is e_P and the last is e_{N-1}. Hence $n_I = P$ and $n_F = N - 1$ in the squared error equation:

$$\mathcal{E}^{\mathrm{C}} = \sum_{n=P}^{N-1} e_n^2$$

The resulting matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is *not* Toeplitz. The AR parameter estimate is not guaranteed to be stable

In the autocorrelation method $n_I = 0$ and $n_F = N + P - 1$. Hence the squared error minimized is:

$$\mathcal{E}^{\mathrm{A}} = \sum_{n=0}^{N+P-1} e_n^2$$

To calculate these error terms requires data before $n = 0$ and after $n = N-1$. These data points are assumed to be zero.

$\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is *Toeplitz*, which means that the efficient Levinson recursion ($O(P^2)$) can be used to solve for $\mathbf{a}$. The parameter estimate is guaranteed to be stable.

Part (c)

The Autocorrelation method:

$$\mathbf{X} = \begin{bmatrix} x_{N_I-1} & x_{N_I-2} & \cdots & x_{N_I-P} \\ x_{N_I} & x_{N_I-1} & \cdots & x_{N_I-P+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N_F-1} & x_{N_F-2} & \cdots & x_{N_F-P} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_{N_I} \\ x_{N_I+1} \\ \vdots \\ x_{N_F} \end{bmatrix}$$

Insert the data to get (remember $N_I = 0$ and $N_F = N - 1 + P = 6$)

$$\mathbf{X} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ -2 & 1 \\ 3 & -2 \\ -4 & 3 \\ 5 & -4 \\ 0 & 5 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \\ -4 \\ 5 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 55 & -40 \\ -40 & 55 \end{bmatrix}, \quad \mathbf{X}^T\mathbf{x} = \begin{bmatrix} -40 \\ 26 \end{bmatrix}$$

Solving $\mathbf{X}^T\mathbf{X}\mathbf{a} = -\mathbf{X}^T\mathbf{x}$ gives $\mathbf{a} = [0.814, 0.119]$. Note that $\mathbf{X}^T\mathbf{X}$ is Toeplitz. The zeros of the transfer function $A(z) = 1 + 0.814z^{-1} + 0.119z^{-2}$ are $z = -0.623$ and $z = -0.191$.