E3.08
ISE3.17

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2007

EEE/ISE PART III/IV: MEng, BEng and ACGI

Corrected Copy

ADVANCED SIGNAL PROCESSING

Monday, 14 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. Petrou, M. Petrou

1) Modern spectrum estimation methods assume that the measured dataset $z[n]$ is generated by some form of model driven by a random noise input. Assume that dataset $z[n]$ was generated by an autoregressive (AR) model of order p ($AR(p)$) driven by white Gaussian noise $w[n]$ with zero mean and variance σ_w^2 ($w[n] \sim \mathcal{N}(0, \sigma^2)$).

a) Write down a general expression for an $AR(p)$ model. [2]

b) Derive the expression for the autocorrelation function of this process and write down the expression for the power spectrum of an $AR(p)$ process. Explain how the power spectrum can be obtained from the functional expression of this process and its autocorrelation function. [8]

c) Figure 1.1 shows the frequency response of an AR process. From the shape of the amplitude spectrum, what is the order of the underlying AR process? Why this cannot be the frequency response of an MA process? [4]

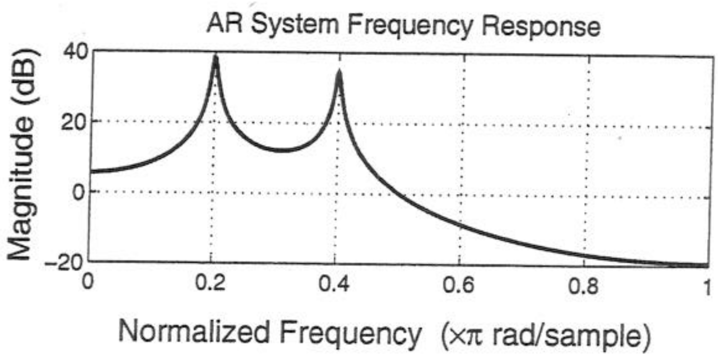


Figure 1.1: Frequency response of an autoregressive process

d) Consider an $AR(2)$ process given by

$$z[n] = 1.5z[n-1] + a_2 z[n-2] + w[n]$$

To preserve stability, should the value of coefficient a_2 be positive or negative? Are the roots of the characteristic polynomial in this case real or complex? Explain. [6]

2) When the probability density function of the data is unknown or cannot be assessed, the minimum variance estimator, even if it exists, cannot be found. One suboptimal approach is to assume the estimator to be linear in the data, unbiased and with minimum variance, such as in the case of Best Linear Unbiased Estimator (BLUE).

a) State the equation for a scalar form of BLUE for the given data

$$\mathbf{x} = \{x[0], x[1], \ldots, x[N-1]\}$$

for which the probability density function (pdf) $p(\mathbf{x}, \theta)$ depends on the unknown parameter θ. [2]

i) Write down the equations describing the constraints of BLUE (linear and unbiased). Explain in your own words the need for these constraints. [6]

ii) The power output of a wind turbine P_{WT} is proportional to the cube of the wind speed v, that is

$$P_{WT} \approx kv^3$$

where k is a constant. Is the BLUE estimator of the average power based on v unbiased? Explain a possible transformation of the wind speed which guarantees unbiased estimation. [4]

b) Consider the case of BLUE for a vector parameter.

i) Write down the expression for the BLUE estimate of the unknown vector parameter $\boldsymbol{\Theta} = [\theta_1, \ldots, \theta_p]^T$ and the variance of such an estimator (Gauss–Markov theorem). [4]

ii) Prove that BLUE is identical to the weighted least squares estimator. Recall that the weighted least squares estimator is found by minimising the criterion

$$J = (\mathbf{x} - \mathbf{H}\boldsymbol{\Theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\Theta})$$

(Hint: To prove that $\boldsymbol{\Theta}$ that minimises J is BLUE, apply the method of least squares to J, based on $\frac{\partial J}{\partial \boldsymbol{\Theta}} = 0$.) [4]

3) A random variable y is estimated in terms of an observation of another random variable x. The problem generally arises when y cannot be directly observed or measured so a related random variable is measured and used to estimate y. The goal is to find the best estimate of y in terms of x.

In the linear mean square estimation, the estimator is constrained to be of the form

$$\hat{y} = ax + b$$

The estimation error is $e = y - \hat{y}$ and the goal is to find the values for a and b that minimize the mean square error

$$J = E\left\{(y - \hat{y})^2\right\} = E\left\{(y - ax - b)^2\right\}$$

a) Solve this linear mean square estimation problem and find the values for a and b. [8]

b) Using the result from a) derive the minimum mean square error J_{min}. [3]

c) What are the advantages of using such an estimator? Does this method require the knowledge of the probability density function? [3]

d) What is the principal difference between this estimator and the standard adaptive finite impulse response (FIR) filter trained by the least mean square (LMS) algorithm? Comment on the character of the cost function to be miminised and the mode of processing (block, sequential). [6]

4) a) An unbiased estimate of the autocorrelation function can be calculated from

$$\hat{r}_{xx}(\tau) = \frac{1}{N-|\tau|} \sum_{k=0}^{N-|\tau|-1} x[k]x[k+\tau], \quad \tau = -(N-1), \ldots, -1, 0, 1, \ldots, (N-1)$$

What is the length of the so obtained autocorrelation function? Explain what is happening for large $|\tau|$. [4]

b) The first ten samples of the normalised autocorrelation function $\rho(k) = r_{xx}(k)/r_{xx}(0)$ of a random process are shown in Figure 3.1. If the underlying process is $AR(p)$, from Figure 3.1 state the order p of such a process and the value of its parameter(s). Write down the exact expression for the power spectrum of this process. Explain. [6]

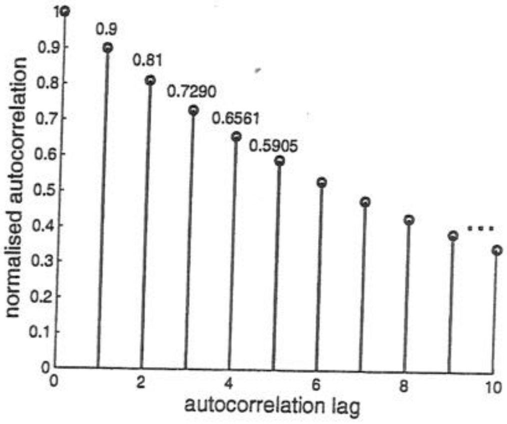


Figure 3.1: Normalised autocorrelation function of a random process

c) Let $x[n]$ be a process that is generated according to the difference equation

$$x[n+1] = a_1 x[n] + w[n]$$

where a_1 is a parameter, and $w[n] \sim \mathcal{N}(0,1)$. An adaptive finite impulse response (FIR) filter is used for the prediction of this process.

i) Derive the expression for the Least Mean Square (LMS) update of a single coefficient adaptive FIR filter, based on the above equation and the result from b). [6]

ii) State the bound on the step size which ensures the convergence of such a filter. [2]

iii) What is the minimum mean square error achievable by using this FIR predictor? [2]

5) Consider the linear optimum filtering problem (Wiener filter). The input–output relationship of this filter is described by

$$y[n] = \sum_{k=1}^{p} w_k x[n-k]$$

where $\mathbf{w}_{opt} = [w_1, w_2, \ldots, w_p]^T$ are filter coefficients and $\{x[0], x[1], \ldots, x[N-1]\}$ the input data. Let $d[n]$ denote the desired response of this filter and $e[n] = d[n] - y[n]$ the instantaneous output error.

a) Derive the optimum set of coefficients for which the mean squared error is minimum (Wiener filter). [4]

b) If the weights of this filter assume a time varying form, explain in your own words the principle behind the method of steepest descent. [6]

c) Draw the block diagram of the system identification configuration of adaptive filtering. Explain the operation of this adaptive filtering scheme. What are the applications of system identification? [4]

i) The cost function for a stochastic gradient type of adaptive filter is given by

$$J[n] = |e[n]|$$

Derive the corresponding learning algorithm for an adaptive FIR filter of length L.

(Hint: The weight update is calculated based on $\nabla_{\mathbf{w}} J[n]_{|\mathbf{w}=\mathbf{w}(k)}$. The derivative of $|e[n]|$ is $sign(e[n]) = \frac{e[n]}{|e[n]|}$. [4]

ii) The sign–sign algorithm is used to train the adaptive filter in the system identification configuration. Write down the expression for the update of such a filter. Explain the benefits and drawbacks as compared to the standard LMS algorithm. [2]

E3.08
ISE 3.17

ADVANCED SIGNAL PROCESSING

2007

1/9 10

Solutions:

1) a) [**bookwork**]
An autoregressive (AR) process of order p, that is $AR(p)$ is given by $z[n] = \sum_{i=1}^{p} a_i z[n-i) + w[n]$
where $z[n]$ is the output of the model and $w[n]$ are samples of zero mean white Gaussian noise with variance σ^2 ($w[n] \sim \mathcal{N}(0, \sigma^2)$).

b) [**bookwork**] Since the autocorrelation function is a function of correlation lag k, we need to calculate $E\{z[n]z[n-k]\}$. TO achieve this, first evaluate the product

$$\begin{aligned} z[n-k]z[n] = a_1 z[n-k]z[n-1] + a_2 z[n-k]z[n-2] + \cdots \\ + a_p z[n-k]z[n-p] + z[n-k]w[n] \end{aligned}$$

Notice that $E\{z[n-k]w[n]\}$ vanishes when $k > 0$, since the driving WGN is not correlated with $z[n]$ for $k > 0$. For the correlation function function we therefore have

$$\begin{aligned} r_{zz}(k) &= a_1 r_{zz}(k-1) + a_2 r_{zz}(k-2) + \cdots + a_p r_{zz}(k-p) \quad k > 0 \quad \textit{and} \\ r_{zz}(0) &= a_1 r_{zz}(1) + a_2 r_{zz}(2) + \cdots + a_p r_{zz}(p) + \sigma_w^2 \quad \text{for k} = 0 \end{aligned}$$

The general expression for the power spectrum of ARMA models is (follows by applying the z transform to the time domain expression for ARMA models)

$$P_z(z) = \sigma_w^2 \frac{B_q(z)B_q(z^{-1})}{A_p(z)A_p(z^{-1})}$$

From this expression and for $B = 1$ we obtain the expression for the power spectrum of an $AR(p)$ process

$$P_{zz}(f) = \frac{2\sigma_w^2}{\left|1 - a_1 e^{-\jmath 2\pi f - \cdots - a_p e^{-\jmath 2\pi p f}}\right|^2} \qquad 0 \le f \le 1/2$$

c) [**bookwork, coursework and intuituve reasoning**] This is an $AR(4)$ model, since the spectrum has two peaks. These peaks are generated by 2 conjugate complex pairs of poles. This cannot be the frequency response of an MA spectrum since there are no finite zeros in the spectrum.

d) [**new example and intuitive reasoning**]
From the stability conditions of AR processes $a_1 + a_2 < 1$, hence $-1 < a_2 < 1$. In this case, since $a_1 > 0$ this means that $a_2 < 0$ and from the second condition also $a_2 > -1$, therefore when $a_1 > 0$, we have $-1 < a_2 < 0$.
Also, from the stability triangle this means that the roots of the characterisic polynomial are complex.

2
$\frac{}{}$ 8 16

2) a)[**bookwork**]
The BLUE estimator is restricted to have the form

$$\theta = \sum_{n=1}^{N-1} a_n x[n]$$

i) [**bookwork**] To determine the BLUE we **constrain** $\hat{\theta}$ to be **linear and unbiased**, and then find the $a_n s$ to minimise the variance. Linearity

$$\theta = \sum_{n=1}^{N-1} a_n x[n]$$

Unbiased constraint

$$E\{\hat{\theta}\} = \sum_{n=1}^{N-1} a_n E\{x[n]\} = \theta \qquad \Rightarrow \qquad \underline{\mathbf{a}}^T \underline{\mathbf{s}} = 1$$

with

$$\underline{\mathbf{s}} = [s[0]\ s[1], \ldots, s[N-1]]^T$$

In other words, in order to satisfy the unbiased constraint for the estimate $\hat{\theta}$, $E\{x[n]\}$ must be linear in θ, or

$$E\{x[n]\} = s[n]\theta$$

ii) [**new example and bookwork**] The MVU estimator P_{WT} is nonlinear in the data v^3. Forcing the estimator to be linear is guaranteed to yield biased estimation. One way round it is to introduce a new variable $z = v^3$, so P_{WT} is now linear in z.

b)
i) [**bookwork**]
Given

$$\underline{\mathbf{x}} = H\ \underline{\hat{\theta}} + \underline{\mathbf{w}}$$

with $\underline{w}$ having zero mean and covariance C, otherwise arbitrary PDF.

BLUE:-

$$\underline{\hat{\theta}} = \left(H^T\ C^{-1}\ H\right)^{-1}\ H^{-1}\ C^{-1}\,\underline{\mathbf{x}}$$

and the minimum variance of θ_i is

$$var(\hat{\theta}_i) = \left[(H^T C^{-1} H)^{-1}\right]_{ii}$$

7/10

with covariance matrix of $\hat{\boldsymbol{\theta}}$

$$C_{\underline{\hat{\theta}}} = \left(H^T \, C^{-1} \, H \right)^{-1}$$

ii) [**new example**] In least squares estimation, we need to find the gradient of the cost function with respect to the unknown parameter vector, and set to zero to find the LS estimates of the coefficients

$$\begin{aligned} \frac{\partial J}{\partial \Theta} &= -2\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} + 2\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\Theta = 0 \\ \Rightarrow \quad \Theta &= \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} \end{aligned}$$

This proves that in this case the BLUE estimator is identical to the weighted LS estimator.

$\frac{4}{16}$

3)a) [**application of bookwork**]
Solving the linear mean square estimation problem starts with differentiating J with respect to a and b and setting the derivatives to zero, as

$$\frac{\partial J}{\partial a} = -2E\{(y - ax - b)x\} = -2E\{xy\} + 2aE\{x^2\} + 2bm_x = 0$$
$$\frac{\partial J}{\partial b} = -2E\{y - ax - b\} = -2m_y + 2am_x + 2b = 0$$

It then follows that

$$a = \frac{E\{xy\} - m_x m_y}{\sigma_x^2}$$
$$b = \frac{E\{x^2\}m_y - E\{xy\}m_x}{\sigma_x^2}$$

Since

$$E\{xy\} = a\sigma_x^2 + m_x m_y$$

we have

$$a = \rho_{xy}\frac{\sigma_y}{\sigma_x}$$
$$b = m_y - am_x$$

b) [**new example**]
Using the values for a and b from part a), the minimum mean square error can be evaluated as

$$E\left\{(y - \hat{y})^2\right\} = \sigma_y^2 - a^2\sigma_x^2 = \sigma_y^2(1 - \rho_{xy}^2)$$

Notice that if x and y are uncorrelated, then $a = 0$ and $b = E\{y\}$.

c) [**bookwork**]
The advantages of using a linear mean square estimator are:
- The parameters a and b depend only on the second order moments of x and y and not on the joint density functions.
- The equations for solving for a and b are linear, hence the computational complexity is low.
- For Gaussian random variables, the optimum mean square estimate is linear.

d) [**bookwork and intuitive reasoning**]
The linear mean square estimate assumes fixed coefficients a and b and the problem is solved in a block fashion, that is, taking into account simultaneously all the available data.
Linear FIR filter, trained by the LMS algorithm, on the other hand deals only

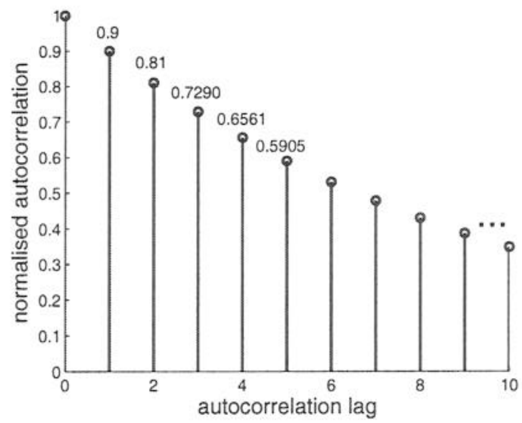
5/10

with a portion of data which is present in its tap inputs, the estimation is recursive, and the coefficients are adaptive.
Least mean squares algorithms minimize a deterministic cost function (sum of squared errors on a block of data), whereas stochastic gradinet algorithms minimise a stochastic cost function (estimation of the instantaneous error).

4) [**coursework solutions, bookwork**]

a) The length L of this ACF is $2N-1$. For large τ there are only a few non–zero elements in the sum and the ACF estimate is not reliable. This is why these estimates are very noisy (see your coursework)

b) [**new example, coursework**]
This is the ACF of an $AR(1)$ process $z[n] = a_1 z[n-1]$, where $a_1 = 0.9$. This is obvious from the expression for ACF of $AR(1)$ processes, $\rho(k) = a_1^k$, and the values of ρ from Figure 3.1. The normalised ACF for $k = 0$, that is $\rho(0) = 1$



and the subsequent ACF coefficients are $\rho(1) = 0.9, \rho(2) = 0.81, \ldots, \rho(k) = 0.9^k$.

The exact expression for the power spectrum of this AR(1) process is (coursework, bookwork)

$$P_{xx}(f) = \frac{1}{|1 + 0.9e^{-\jmath 1\pi f}|^2}$$

c) i) [**application of bookwork and coursework**]
i) Here, we need to find adaptively the value of the unknown coefficient a which generates this process. This will be achieved if the instantaneous output error of the adaptive filter is white (in this case the "error" is $w[n]$), and we minimise the instantaneous estimate of the error power $E(m) = \frac{1}{2}e^2(n)$. Following the standard LMS update

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

where $\mathbf{w}$ are filter coefficients, μ learning rate, $e(n)$ is the instantaneous output error and $\mathbf{x}(n)$ the input signal in filter memory, we have

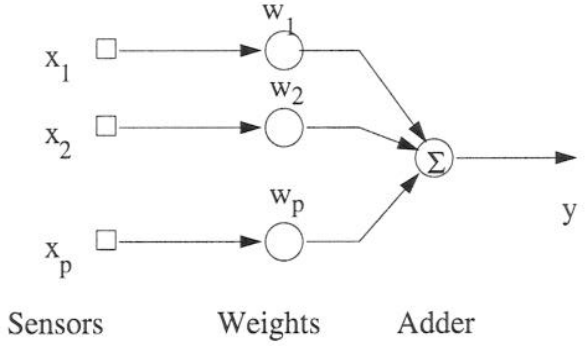
$$\begin{aligned} \hat{x}[n+1] &= a_1[n]\hat{x}[n] \\ a_1[n+1] &= a_1[n] + \mu e[n]x[n] \end{aligned}$$

ii) and iii) [**applied bookwork**] Same as for standard LMS with length $N = 1$.

5)
a) [**bookwork**]

Consider a set of p sensors at different points in space

Let $x_1, x_2, \ldots, x_p$ be the individual signals from sensors



- The signals are applied to a corresponding set of weights $w_1, \ldots, w_p$
- The weighted signals are them summed to produce the output y

$$y(n) = \sum_{i=1}^{p} w_i x_i = \mathbf{x}^T \mathbf{w}$$

- **Requirement:** To determinde the optimum setting of weights $\mathbf{w} = [w_1, \ldots, w_p]^T$ so as to minimize the difference between the system output and some desired response d in the mean square sense.

The solution to this fundamental problem lies in the Wiener–Hopf equations

The system from the Figure can be seen as a *spatial* filter.

The input-output relation of the filter is given by

$$y = \sum_{k=1}^{p} w_k x_k$$

Let d denote the *desired response* or *target output* for the filter. Then the *error signal* is

$$e = d - y$$

As *performance measure* or *cost function*, we introduce the *mean squared error* defined as

$$J = \frac{1}{2} E\{e^2\}$$

8/16

(the factor 1/2 included for convenience)

Statement of the *linear optimum filtering problem:*

Determine the optimum set of weights $w_{01}, \ldots, w_{0p}$ for which the mean squared error J is minimum.

The solution to this problem is known as the **Wiener Filter**

$$J = \frac{1}{2}E\{d^2\} - E\left\{\sum_{k=1}^{p} w_k x_k d\right\} + \frac{1}{2}E\left\{\sum_{j=1}^{p}\sum_{k=1}^{p} w_j w_k x_j x_k\right\}$$

where the double summation represents the square of a summation.

$$J = \frac{1}{2}E\{d^2\} - \sum_{k=1}^{p} w_k E\{x_k d\} + \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p} w_j w_k E\{x_j x_k\}$$

Notation: $r_d = E\{d^2\}$
$r_{dx}(k) = E\{dx_k\}, \qquad k = 1, 2, \ldots, p$
$r_x(j,k) = E\{x_j x_k\}, \qquad j, k = 1, 2, \ldots, p$
Slot back into J to yield

$$J = \frac{1}{2}r_d - \sum_{k=1}^{p} w_k r_{dx}(k) + \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p} w_j w_k r_x(j,k)$$

Def: A multidimensional plot of the cost function J verus the weights (free parameters) $w_1, \ldots, w_p$ consitutes the **error performance surface** or simply the **error surface** of the filter.

The error surface is bowl–shaped with a well–defined bottom or global minimum point. It is precisely at this point where the spatil filter from the FIgure is optimum in the sense that the mean squared error attins its minimum value J_{min}.

To determinde the optimum weights, follow the least squares approach

$$\nabla_{w_k} J = \frac{\partial J}{\partial w_k}, \qquad k = 1, \ldots, p$$

Differentiate wrt to w_k

$$\nabla_{w_k} J = -r_{dx}(k) + \sum_{j=1}^{p} w_j r_x(j,k)$$

and set to zero

$$\nabla_{w_k} J = 0, \qquad k = 1, 2, \ldots, p$$

Let w_{0k} denote the optimum setting of weight w_k. Then the optimum weights are determined by the following set of simultaneous equations

$$\sum_{j=1}^{p} w_{0j} r_x(j,k) = r_{xd}(k), \qquad k = 1, 2, \ldots, p$$

or in a compact form

$$\mathbf{W}_{opt} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{dx}$$

This system of equations is known as the **Wiener-Hopf** equations and the filter whose weights satisfy the Wiener Hopf equations is called a **Wiener** filter.

Notice, this is a block filter, operating on the whole set of data

b) [**bookwork**]

The idea is to replace the block estimate from the Wiener filter with a recursive estimate on a much shorter data length. This allows for a sequential solution of this block filtering problem, which facilitates the use of short filters. However this way we introduce an error in the estimation.

Problem with the Wiener filter: Matrix inversion of the $p \times p$ matrix $\mathbf{R}$.

We may avoid the need for matrix inversion

by using the method of steepest descent.

Difference from Wiener Filter: The weights have a **time–varying** form,

they are adjusted in an **iterative** fashion allong the error surface.

The gradient of the error surface of the filter wrt the weights takes on a *time varying* form

$$\nabla_{w_k} J(n) = -r_{dx}(k) + \sum_{j=1}^{p} w_j(n) r_x(j,k)$$

(indices j, k refer to locations of different sensors in space, whereas the index n refers to iteration number).

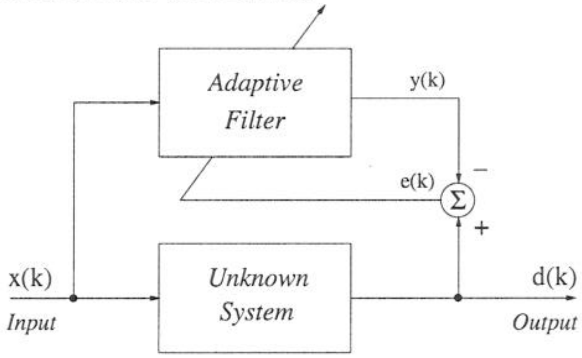
According to the method of steepest descent, tje adjustment applied to the weight $w_k(n)$ at iteration n is defined by

$$\Delta w_k(n) = -\eta \nabla_{w_k} J(n), \qquad k = 1, 2, \ldots, p$$

10/16

where η is a positive constant called the **learning rate** parameter (step size).

c) [**new example based on coursework**]
Both the unknown system and the adaptive filter have the sam input, and their outputs are compared to produce the instantaneous output error $e(k)$ which is then used to update the filter coefficients.



Applications include acoustic echo cancellation, plant modellig.

i) This is the cost function of a sign error algorithm. Perform $\frac{\partial J[n]}{\partial \mathbf{w}[n]}$ to obtain

$$\frac{\partial J[n]}{\partial \mathbf{w}[n]} = \frac{|e[n]|}{\partial \mathbf{w}[n]} = \frac{|e[n]|}{\partial e[n]} \frac{\partial e[n]}{\partial y[n]} \frac{\partial y[n]}{\partial \mathbf{w}[n]} = sign(e[n])(-1)\mathbf{x}[n]$$

Now from

$$\begin{aligned}
\mathbf{w}[n+1] &= \mathbf{w}[n] - \mu \nabla_{\mathbf{w}} J[n]_{|\mathbf{w}=\mathbf{w}[n]} \\
&\mathit{we\ \ have\ \ the\ \ update} \\
\mathbf{w}[n+1] &= \mathbf{w}[n] + \mu sign(e[n])\mathbf{x}[n]
\end{aligned}$$

ii) In very much the same way as above, apply the sign operator to both the error $e[n]$ and the regressor vector $\mathbf{x}[n]$ to obtain

$$w[n+1] = w[n] + \mu sign(e[n]) sign\left(\mathbf{x}[n]\right)$$

This algorithm requires lower computational complexity, but cannot achieve the same steady state accuracy as standard LMS.