ADAPTIVE DIGITAL FILTERS: PART I

ADAPTIVE NOISE CANCELLERS

When a random signal is corrupted by an additive random noise with overlapping signal and noise spectra, Wiener filtering technique is found to be the optimum for stationary processes. The design of these optimal filters requires *a priori* knowledge of both the signal and the noise. Adaptive filters, on the other hand, have the ability to adjust their own parameters automatically to reach the performance of optimal filters, while their design requires little or no *a priori* knowledge of signal and noise statistics.

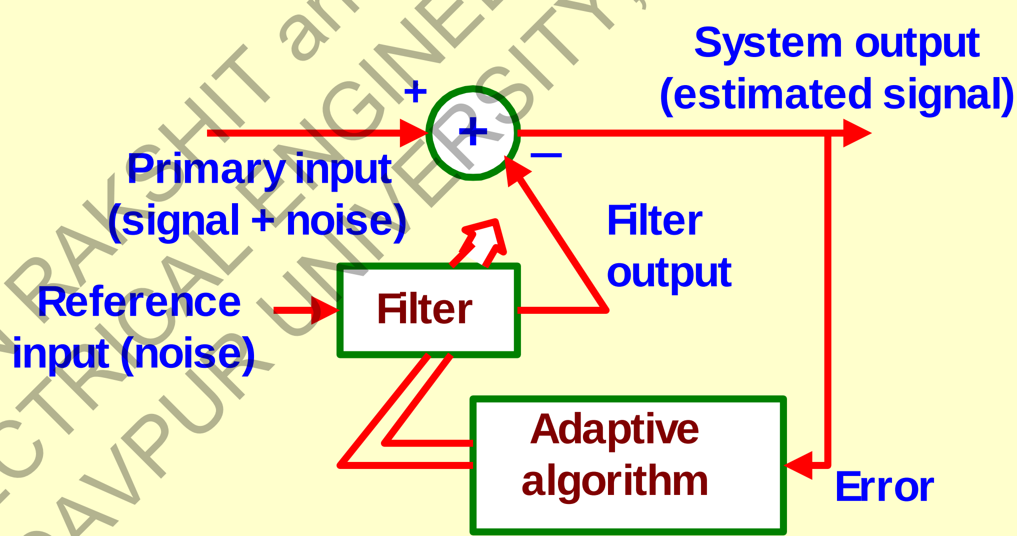


Fig. 1: The basic adaptive noise canceller

In Fig.1, a reference input from the noise field is filtered and subtracted from the primary input containing both signal and noise.

The difference between the primary input and the filter output adjusts the filter parameters through an adaptive algorithm so that the primary noise is attenuated or eliminated by cancellation.

Let the primary input sample be P_n at the nth instant and it is

$$P_n = S_n + N_n \qquad (1)$$

where S_n is the signal sample at the nth instant and N_n is the additive noise sample at the nth instant. If the reference input at the nth instant be R_n and is assumed to be correlated to the noise but uncorrelated to the signal, then

$$\left.\begin{aligned} & E\left[R_n N_k\right] \neq 0 \\ \& \quad & E\left[R_n S_k\right] = 0 \end{aligned}\right\} \qquad (2)$$

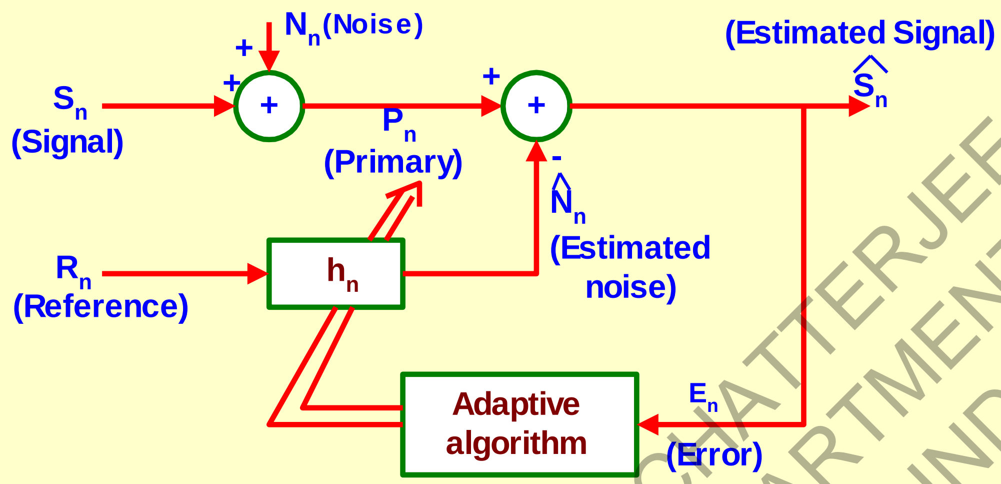


Fig. 2: Adaptive noise canceller with correlated noise in the reference input

In fig. 2, the reference input R_n is filtered through the filter H(z) to give the estimated noise sample $\hat{N}_n$ which is subtracted from the primary sample to yield the signal estimate $\hat{S}_n$. The resulting error E_n which happens to be the estimated signal itself, ultimately adjusts the parameters of the filter system function H(z).

The adaptive algorithm adjusts the filter parameters so that the error converges to a minimum in the **mean – square sense**. Then the adaptive filter becomes an equivalent to the Wiener filter. The filter impulse response h_{wj} at this optimal condition is such that

$$h_{wj} * R_{RR}(j) = R_{RP}(j) \qquad (3)$$

where $R_{RR}(j)$ is the discrete autocorrelation function of the reference input, and $R_{RP}(j)$ is the discrete cross-correlation function between the reference and the primary inputs. By applying **Wiener – Khintchine theorem**

$$S_{RP}(z) = S_{RR}(z) H_w(z) \qquad (4)$$

where $S_{RP}(z)$ is the cross spectral density between the reference and primary inputs, $S_{RR}(z)$ is the power spectral density of the reference input and $H_w(z)$ is the system function of the optimal filter.

Now,

$$S_{RP}(z) = S_{RS}(z) + S_{RN}(z) \qquad (5)$$

Assuming that the original signal and the reference input are completely uncorrelated,

$$S_{RS}(z) = 0$$

hence, $S_{RP}(z) = S_{RN}(z)$ (6)

Therefore, from (4) and (6),

$$H_w(z) = S_{RN}(z) \big/ S_{RR}(z) \qquad (7)$$

This form of Wiener solution is unconstrained, because here the impulse response h_{wj} may be causal or non-causal and of finite or infinite duration.

For physical realization, the impulse response is normally truncated and delayed and realized with FIR filter structures.

Adaptive Cancellation of Noise without an External Reference Source:

In some applications no external reference input, free from the signal, is available for adaptive processing. In such cases, the reference has to be derived from the primary input itself. Let a reference filter F(z) be included between the primary and reference inputs so that the filter provides, as output, a reference R_n which is highly correlated to noise N_n but is almost uncorrelated to the signal S_n as shown in Fig. 3.

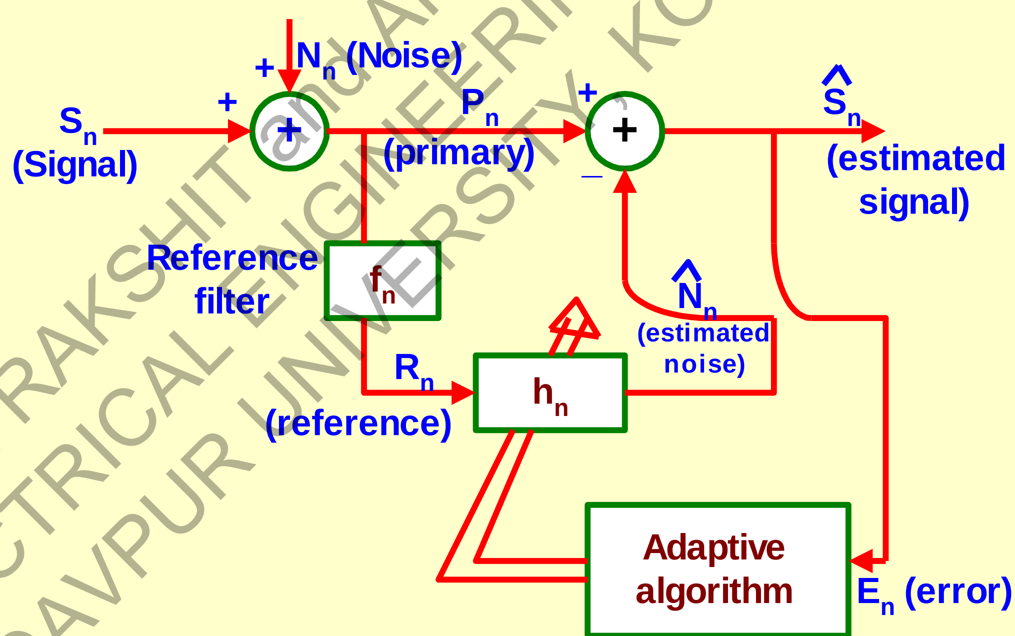


Fig.3 : Adaptive noise canceller with reference derived from the primary input

The reference filter F(z) may be replaced by a **bulk delay stage** when the signal is random but the noise is periodic (e.g. sinusoidal interference). This delay decorrelates the signal in the reference path and so may be called a **decorrelation delay**. Due to its periodic nature the noise remains correlated in the reference path.

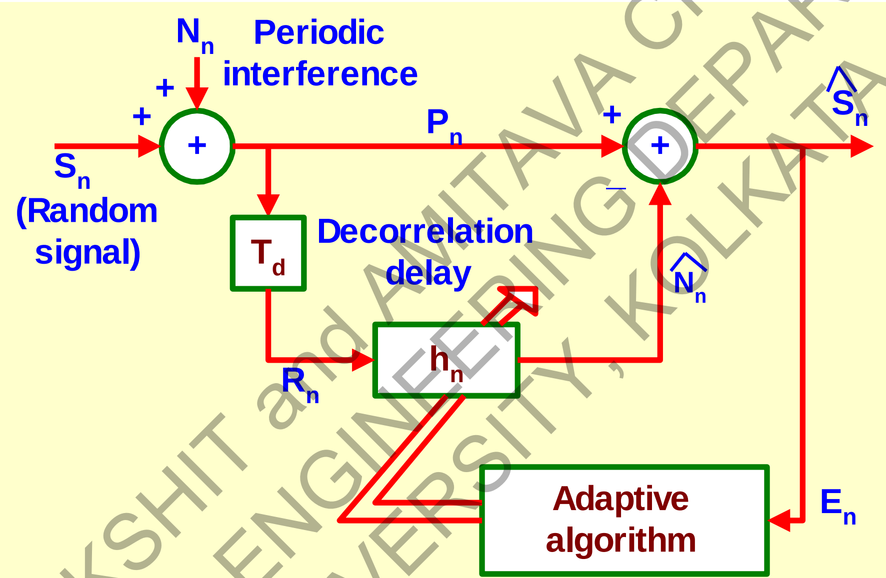


Fig. 4: Adaptive noise canceller for periodic interference without an external reference source

In case, the signal contains correlated components (rather than being purely random), improper selection of decorrelation delay may lead to appreciable leakage of correlated signal in the reference path. As a result, the filter not only cancels the interference but also distorts, in general, the estimated signal at the noise canceller output.

THE ADAPTIVE DIGITAL FILTER

Let an M-weight FIR adaptive filter be realized from an M-tapped delay line as shown in Fig. 5.

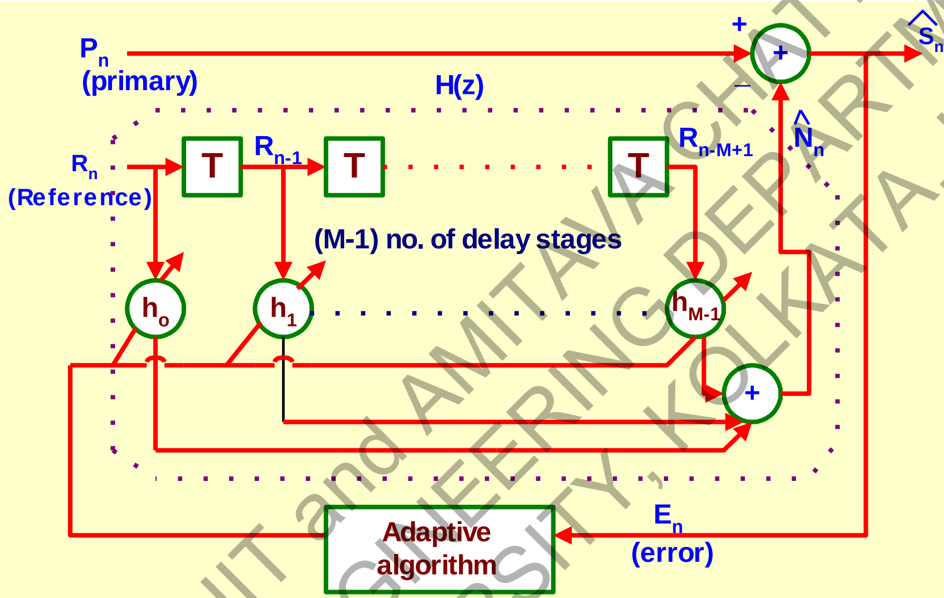


Fig. 5: The Adaptive Filter

The adaptive filter weights, h_k, k = 0, 1, …………., M – 1 are updated by the adaptive algorithm at the input sampling rate $1/T$.

The filter output $\hat{N}_n$ can be expressed as

$$\hat{N}_n = \sum_{k=0}^{M-1} h_k R_{n-k} \qquad (12)$$

and the error E_n is the same as the estimated signal $\hat{S}_n$ and it is

$$\therefore E_n = \hat{S}_n = P_n - \hat{N}_n \qquad (13)$$

In terms of matrix notations,

$$\hat{N}_n = H^T R_n = R_n^T . H \qquad (14)$$

where $H^T = [h_o, h_1, \ldots\ldots\ldots h_{M-1}]$, the transposed weight vector matrix,

and $R_n^T = [R_n, R_{n-1}, \ldots\ldots\ldots R_{n-M+1}]$, the transposed reference matrix.

The error at the nth instant is

$$\begin{aligned} E_n &= P_n - \hat{N}_n \\ &= P_n - H^T R_n = P_n - R_n^T H \end{aligned} \qquad (15)$$

The square of this error is

$$E_n^2 = P_n^2 - 2P_n R_n^T H + H^T R_n R_n^T H \qquad (16)$$

The mean square error ξ , i.e., the expected value of E_n^2 is

$$\xi = E\left[E_n^2\right] = E\left[P_n^2\right] - 2E\left[P_n R_n^T\right]H + H^T E\left[R_n R_n^T\right]H$$

or, $$\xi = E\left[P_n^2\right] - 2P^T H + H^T RH \qquad (17)$$

where $P^T = [R_{PR}(0) \ldots\ldots\ldots\ldots\ldots R_{PR}(M-1)]$, the cross-correlation matrix between primary and reference

and $$R = \begin{bmatrix} R_{RR}(0) & \ldots\ldots\ldots\ldots & R_{RR}(M-1) \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ R_{RR}(M-1) & \ldots\ldots\ldots\ldots & R_{RR}(0) \end{bmatrix}$$, the reference correlation matrix or $$R = E\begin{bmatrix} R_n R_n & R_n R_{n-1} \ldots\ldots\ldots\ldots\ldots R_n R_{n-M+1} \\ R_{n-1} R_n & R_{n-1} R_{n-1} \qquad\qquad : \\ : & \ddots \qquad : \\ : & \ddots \quad : \\ R_{n-M+1} R_n & \ldots\ldots\ldots\ldots\ldots \quad R_{n-M+1} R_{n-M+1} \end{bmatrix}$$

It may be observed from relation (17) that the mean square error (mse) is a quadratic function of weights, a (M+1) dimensional paraboloid surface. The optimum filtering corresponds to the bottom of the bowl. The adaptive algorithm seeks the bottom of the bowl (shown in Fig. 6).

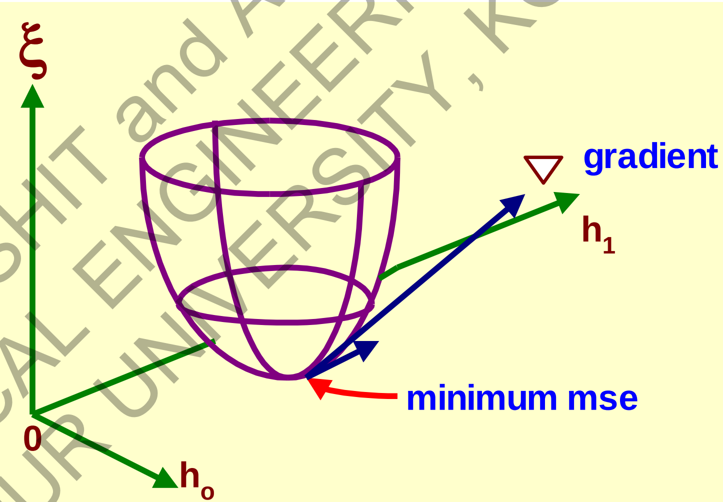


Fig.6: Mean Square error surface for two weights

The adaptive algorithm utilizes the **method of steepest descent** in seeking the minimum mse. The gradient at any point on the error surface may be obtained by differentiating relation (17) with respect to weight vector. The gradient vector is

$$\nabla = \begin{Bmatrix} \dfrac{\partial \xi}{\partial h_o} \\ \vdots \\ \vdots \\ \dfrac{\partial \xi}{\partial h_{M-1}} \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial E\left[E_n^2\right]}{\partial h_o} \\ \vdots \\ \vdots \\ \dfrac{\partial E\left[E_n^2\right]}{\partial h_{M-1}} \end{Bmatrix} = -2P + 2RH \qquad (18)$$

The optimal weight vector H_w, generally called the Wiener weight vector, is obtained by setting the gradient of the mse function to zero, i.e.,

$$0 = -2P + 2RH_w$$

or $$H_w = R^{-1} P \qquad (19)$$

This is the matrix form of the **Wiener-Hopf equation**.

The minimum mse is obtained from relations (17) and (19) as (under optimal condition)

$$\xi_{\min} = E\left[P_n^2\right] - 2P^T H_w + H_w^T R H_w$$

or $$\xi_{min} = E\left[P_n^2\right] - 2P^T H_w + H_w^T P \qquad \left(\text{since, } P = RH_w\right)$$

or $$\xi_{\min} = E\left[P_n^2\right] - P^T H_w \qquad (20)$$

Now, we can write,

$$\begin{aligned}
\xi &= \xi_{\min} + P^T H_w - 2P^T H + H^T RH \\
&= \xi_{\min} + H_w^T RH_w - 2H_w^T RH + H^T RH \\
&\qquad\qquad \left(\text{since } R \text{ is symmetric, } R = R^T\right) \\
&= \xi_{\min} + H^T RH - H^T RH_w - H_w^T RH + H_w^T RH_w \\
&= \xi_{\min} + \left(H^T - H_w^T\right)(RH - RH_w) \\
\text{or, } \xi &= \xi_{\min} + (H - H_w)^T R(H - H_w)
\end{aligned} \tag{21}$$

Let V be the difference between H and the Wiener solution H_w.

Then $V = (H - H_w)$. (22)

Then,

$$\xi = \xi_{\min} + V^T RV \tag{23}$$

The gradient ∇ can also be expressed by differentiating relation (23) as

gradient matrix

$$\nabla = 2RV \tag{24}$$

The reference autocorrelation matrix being symmetric and positive definite, may be represented as

$$R = Q \wedge Q^{-1} \tag{25}$$

similarity transform

where Q is the orthonormal (unity norms) modal matrix of R and Λ is its diagonal matrix of eigen values.

$$\wedge = \begin{bmatrix} \lambda_o \, 0 \, 0 & \cdots & 0 \\ 0 \, \lambda_1 \, 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 \, 0 \cdots 0 & & \lambda_{M-1} \end{bmatrix} \qquad (26)$$

Each column of Q represents the eigen vectors of R, corresponding to each eigen value λ_n. Also,

$$Q^{-1} = Q^T \qquad (27)$$

as Q is orthogonal (i.e, $Q^TQ = I$) with unity norms. Then relation (23) can be represented as

$$\xi = \xi_{\min} + V^T Q \wedge Q^{-1} V \qquad (28)$$

Let a transformed version of V be

$$V' = Q^{-1}V \quad \text{and} \quad V = QV' \qquad (29)$$

then relation (28) becomes,

$$\xi = \xi_{\min} + V'^T \wedge V' \qquad (30)$$

The primed co-ordinates are therefore the principal axes of the ***mse*** surface (shown in Fig. 7).

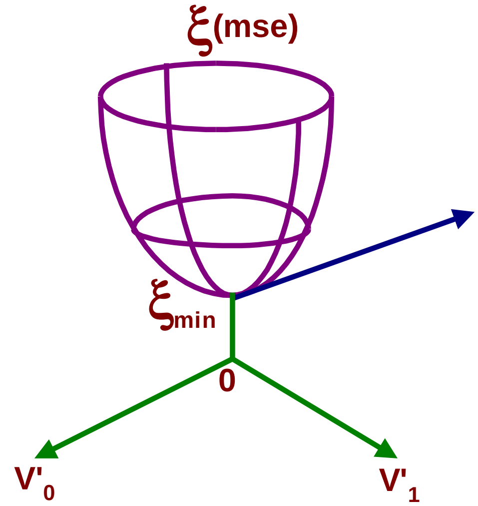


Fig.7: Primed coordinates for two weight error surface

$$V_o' = \left(h_o' - h_{w_0}'\right)$$
$$V_1' = \left(h_1' - h_{w_1}'\right)$$

where h' are the transformed weights.

Here V_0' is the dimension 0 of vector V' and V_1' is the dimension 1 of vector V'.

The **method of steepest descent** makes each change in the weight vector proportional to the negative of the gradient vector at the end of the nth iteration as

$$H_{n+1} = H_n + \mu(-\nabla_n) \quad (31)$$

where μ is the *feedback co-efficient* (a scalar quantity) that controls the stability and the rate of convergence. Each iteration occupies a unit time period T. The gradient at the nth iteration is represented by ∇_n. H_n represents the weight vector matrix at the nth instant. Using relations (24) , (25) and (29), relation (31) becomes,

$$H_{n+1} = H_n - 2\mu R V_n$$

or $(H_{n+1} - H_w) = (H_n - H_w) - 2\mu R V_n$

or $V_{n+1} = V_n - 2\mu R V_n$

$= V_n - 2\mu Q \wedge Q^{-1} V_n$ (32)

or $V'_{n+1} = V'_n - 2\mu \wedge V'_n$ (premultiplying by Q^{-1})

or $V'_{n+1} - (I - 2\mu \wedge) V'_n = 0$

where I is the unity matrix $\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$

Relation (32) is uncoupled and each mode can be solved independently.

The initial condition can be given as,

$$V'_1 = (I - 2\mu \wedge) V'_o$$

where V'_o is the initial value of V'.

(V'_o = vector V' at iteration 0,

V'_1 = vector V' at iteration 1,

V'_2 = vector V' at iteration 2, and so on)

Then, at the next instant,

$$V_2' = (I - 2\mu \wedge)V_1' = (I - 2\mu \wedge)^2 V_o'$$

Then, at the nth instant,

$$V_n' = (I - 2\mu \wedge)^n V_o' \qquad (33)$$

Now, for convergence, it is necessary that, under pth mode, $\left(\underset{k \to \infty}{Lt} V_{p_k}' = 0\right)$, where V_{p_k}' : pth mode of V' at kth iteration.

$$\left|1 - 2\mu\lambda_p\right| < 1$$

$$\text{or, } 0 < \mu\lambda_p < 1$$

$$\text{or, } \frac{1}{\lambda_p} > \mu > 0$$

Then under all modes, $\frac{1}{\lambda_{max}} > \mu > 0$ (34)

where λ_{max} is the largest eigen value of R. Relation (34) is the **necessary condition for convergence**.

From relation (33) it can be seen that the transients in the primed coordinates will be geometric and the geometric ratio of the pth coordinate is

$$r_p = (1 - 2\mu\lambda_p) \qquad (35)$$

where λ_p is the pth eigen value of the correlation matrix R.

An exponential envelope can be fitted to a geometric sequence.

From relations (33) and (35) , for the pth mode,

$$V'_{p_n} = r_p^n V'_{p_o} \tag{36}$$

and from relation (32) and (35) , for the pth mode,

$$V'_{p_n} = r_p V'_{p_{(n-1)}} \tag{37}$$

Relation (37) is represented in Fig. 8 as

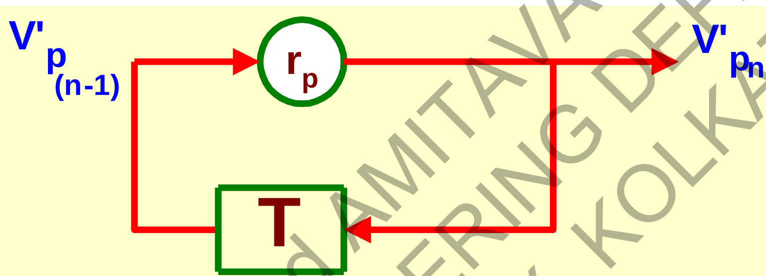


Fig.8

Relation (36) can be rewritten as

$$V'_{p_n} = e^{-\left(nT /_{\tau_p T} \right)} V'_{p_o} \tag{38}$$

Comparing with (36), we get

$r_p^n = e^{- n/_{\tau_p}}$, [time = t = nT]

or, $$r_p = e^{-\frac{1}{\tau_p}} \tag{39}$$

where τ_p is the time constant, expressed in number of iteration cycles, for the pth mode. Fig. 9 shows the exponential envelope represented by relation (38)

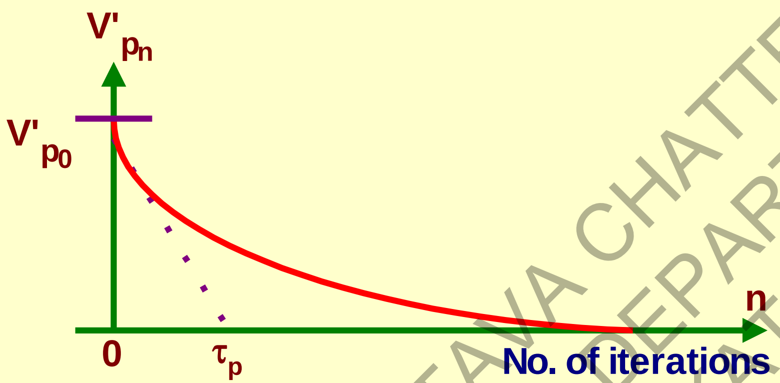


Fig.9

Now, relation (38) can be approximated as

$$r_p = e^{-\frac{1}{\tau_p}} \approx 1 - \frac{1}{\tau_p} + \frac{1}{2!\,\tau_p^2} \ldots\ldots\ldots$$
$$\approx 1 - \frac{1}{\tau_p} \quad \text{, for large } \tau_p \qquad (40)$$

Then from relation (35) and (40),

$$r_p = 1 - 2\mu\lambda_p \approx 1 - \frac{1}{\tau_p}$$
$$\text{or,} \quad \tau_p \approx \frac{1}{2\mu\lambda_p} \qquad (41)$$

for large τ_p

Relation (41) gives the time constant of the pth mode.

The mean square error (mse) at the nth iteration can be expressed from relation (30) as,

$$\xi_n = \xi_{\min} + V_n'^T \wedge V_n' \tag{42}$$

Assuming no noise in the weight vectors during adaptation, the mse can be expressed (from relations (42) and (33)) as,

$$\xi_n = \xi_{\min} + V_o'^T \wedge (I - 2\mu \wedge)^{2n} V_o' \tag{43}$$

When the adaptation process is convergent, then $\underset{n \to \infty}{Lt}\ \xi_n = \xi_{\min}$.

From relation (43), decay in ξ_n, going from ξ_o to ξ_{min} will have a geometric ratio of r_p^2 for the pth mode and it is

$$r_p^2 = (1 - 2\mu\lambda_p)^2 \tag{44}$$

Therefore, the corresponding time constant of decay of mse, under pth mode, is

$$\tau_{p_{mse}} = \frac{\tau_p}{2} \approx \frac{1}{4\mu\lambda_p} \tag{45}$$

for large τ_p

$$\left(\text{since,} \quad r_p^2 = e^{-\frac{2}{\tau_p}} = e^{-\frac{1}{\tau_{p_{mse}}}} \right)$$

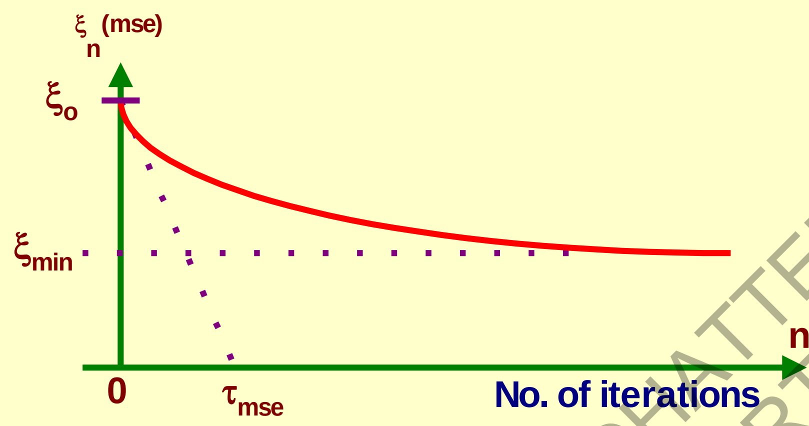


Fig.10: Learning curve

The curve representing the variation of ***mse*** with number of iterations is known as the **learning curve**. Due to noise in the weight vector, actual ***mse*** is generally higher than indicated by relation (43).

Although the learning curve consists of a sum of exponentials, it can be approximated by a single exponential (shown in Fig. 10) whose time constant τ_{mse} is given as

$$\tau_{mse} = \frac{1}{4\mu\lambda_{av}}$$ cycles, where λ_{av} = average of eigen values

$$= \frac{\lambda_o + .. + \lambda_{M-1}}{M} = \frac{trR}{M}$$

Condition (34) is necessary and sufficient for convergence of the **steepest descent method**. However, in practice, the individual eigen values are rarely known. Since trR is the total power input to the weights, trR is a known quantity, and

$$trR = \sum_{k=0}^{M-1} \lambda_k \tag{46}$$

then,

$$\text{trR} \geq \lambda_{\max} \tag{47}$$

as R is positive definite.

Therefore, a sufficient condition for convergence is

$$\frac{1}{trR} > \mu > 0 \tag{48}$$

The Widrow-Hoff LMS Algorithm

The **LMS (least mean square) algorithm** is an implementation of the steepest descent using measured or estimated gradients. The LMS algorithm estimates an instantaneous gradient in a crude but efficient manner by assuming that E_n^2, the square of a single error sample, is an estimate of the mean square error $\mathrm{E}\left[E_n^2\right]$, i.e.

$$\hat{\xi} = \hat{E}\left[E_n^2\right] = E_n^2 \qquad (49)$$

By differentiating E_n^2 with respect to H, we obtain the estimated gradient at the nth iteration (in matrix form), given as,

$$\hat{\nabla}_n = \begin{Bmatrix} \dfrac{\partial E_n^2}{\partial h_o} \\ \vdots \\ \vdots \\ \dfrac{\partial E_n^2}{\partial h_{M-1}} \end{Bmatrix} = 2E_n \begin{Bmatrix} \dfrac{\partial E_n}{\partial h_o} \\ \vdots \\ \vdots \\ \dfrac{\partial E_n}{\partial h_{M-1}} \end{Bmatrix} \qquad (50)$$

From relation (15),

$$E_n = P_n - H^T R_n,$$

thus, $\hat{\nabla}_n = 2E_n(-R_n) = -2E_n R_n$

The kth mode of gradient estimate is $\hat{\nabla}_{k(n)} = -2E_n R_{n-k}$ (51)

This is because,

$$\hat{\nabla}_n = -2E_n \begin{bmatrix} R_n \\ R_{n-1} \\ \vdots \\ R_{n-k} \\ \vdots \\ R_{n-M+1} \end{bmatrix} = \begin{bmatrix} \hat{\nabla}_{0(n)} \\ \hat{\nabla}_{1(n)} \\ \vdots \\ \hat{\nabla}_{k(n)} \\ \vdots \\ \hat{\nabla}_{(M-1)(n)} \end{bmatrix}$$

Putting this value of estimated gradient in relation (31), yields the **Widrow-Hoff LMS algorithm**,

$$\begin{aligned} H_{n+1} &= H_n + \mu\left(-\hat{\nabla}_n\right) \\ &= H_n + 2\mu E_n R_n \end{aligned} \qquad (52)$$

To determine an expression for each weight update individually, we can write,

$$\underbrace{\begin{bmatrix} h_{0(n+1)} \\ h_{1(n+1)} \\ \vdots \\ h_{k(n+1)} \\ \vdots \\ h_{(M-1)(n+1)} \end{bmatrix}}_{H_{n+1}} = \underbrace{\begin{bmatrix} h_{0(n)} \\ h_{1(n)} \\ \vdots \\ h_{k(n)} \\ \vdots \\ h_{(M-1)(n)} \end{bmatrix}}_{H_n} + 2\mu E_n \underbrace{\begin{bmatrix} R_n \\ R_{n-1} \\ \vdots \\ R_{n-k} \\ \vdots \\ R_{n-M+1} \end{bmatrix}}_{R_n}$$

Then, for the kth weight,

$$h_{k(n+1)} = h_{k(n)} + 2\mu E_n R_{n-k} \qquad (53)$$

Fig. 11 shows a schematic representation of the **LMS algorithm**.

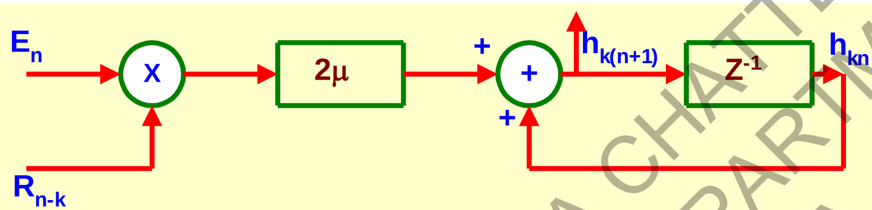


Fig.11: LMS algorithm for the computation of kth weight at the nth instant

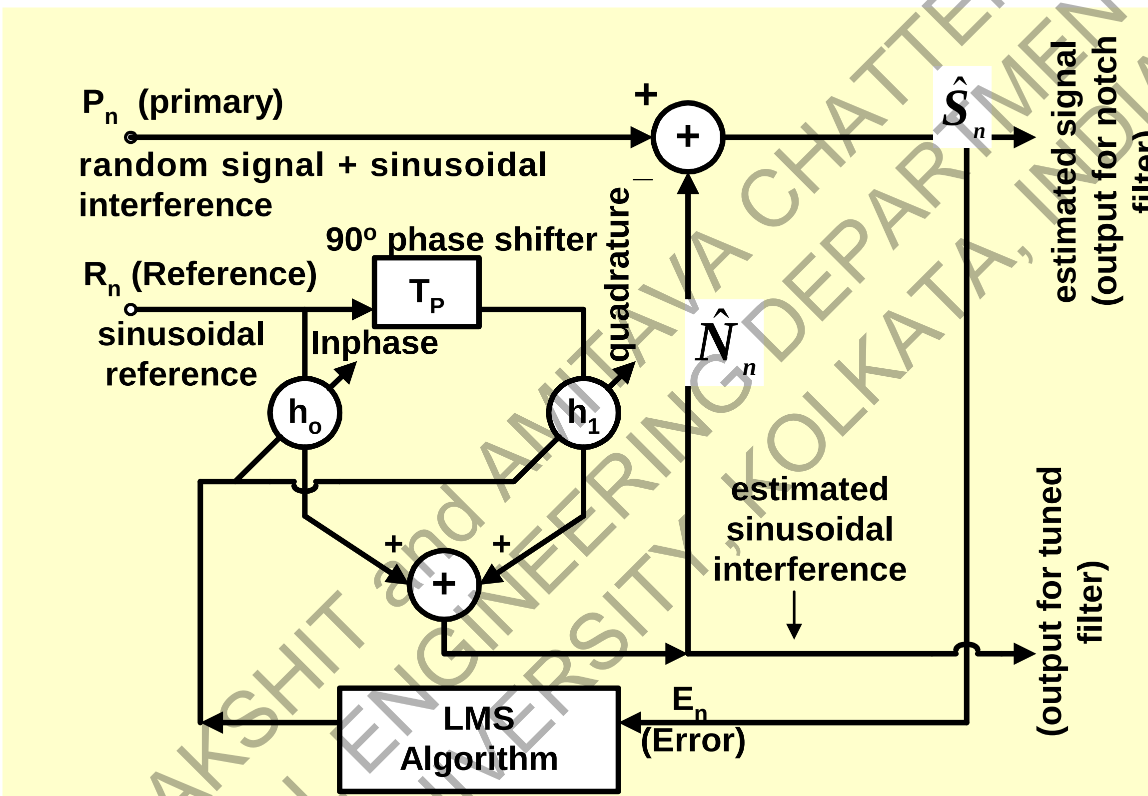
This algorithm is very simple and requires only M number of additions and M number of multiplications per iteration, for the filtering purpose, and M number of additions and M number of multiplications for computation of weights, i.e., a total of 2M number of additions and 2M number of multiplications are required.

The step size μ in the LMS algorithm was originally chosen to be fixed. However, there can be both variations possible i.e. either μ is kept fixed or μ is adapted over iterations. In adaptive filtering problems, it is common to use a fixed μ because of mainly two reasons:

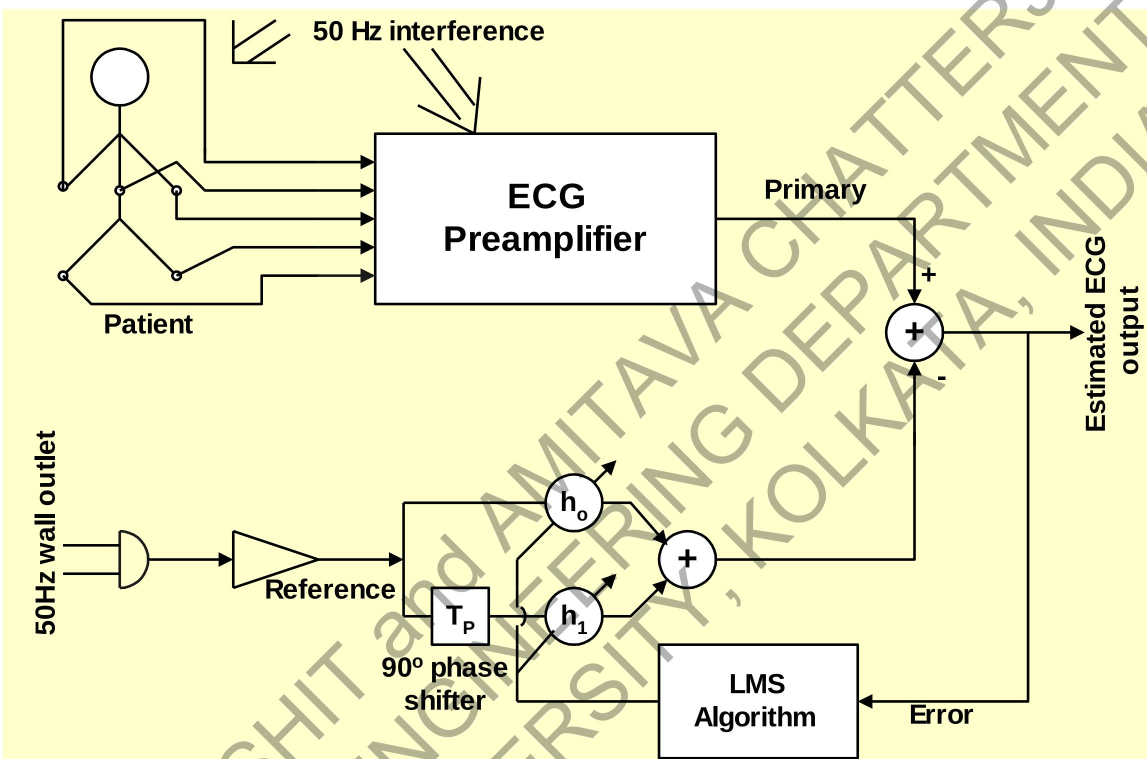
- A fixed-step-size algorithm can be easily implemented in both hardware and software.
- A fixed-step-size is appropriate for tracking time-variant signal statistics, whereas, if μ is adapted over iterations and if $\mu_n \rightarrow 0$ as $n \rightarrow \propto$, adaptation to signal variations cannot occur.

OTHER APPLICATIONS OF ADAPTIVE DIGITAL FILTERS

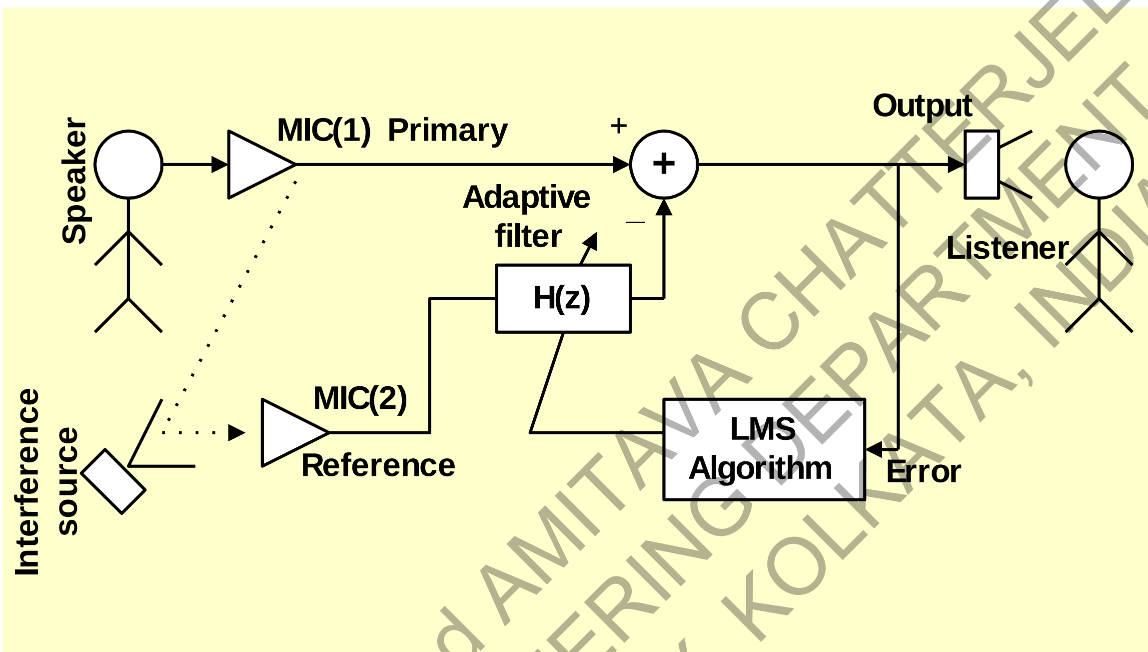
a) *The adaptive noise canceller as a notch filter:*



Application of single frequency adaptive noise cancellation in on-line ECG processing

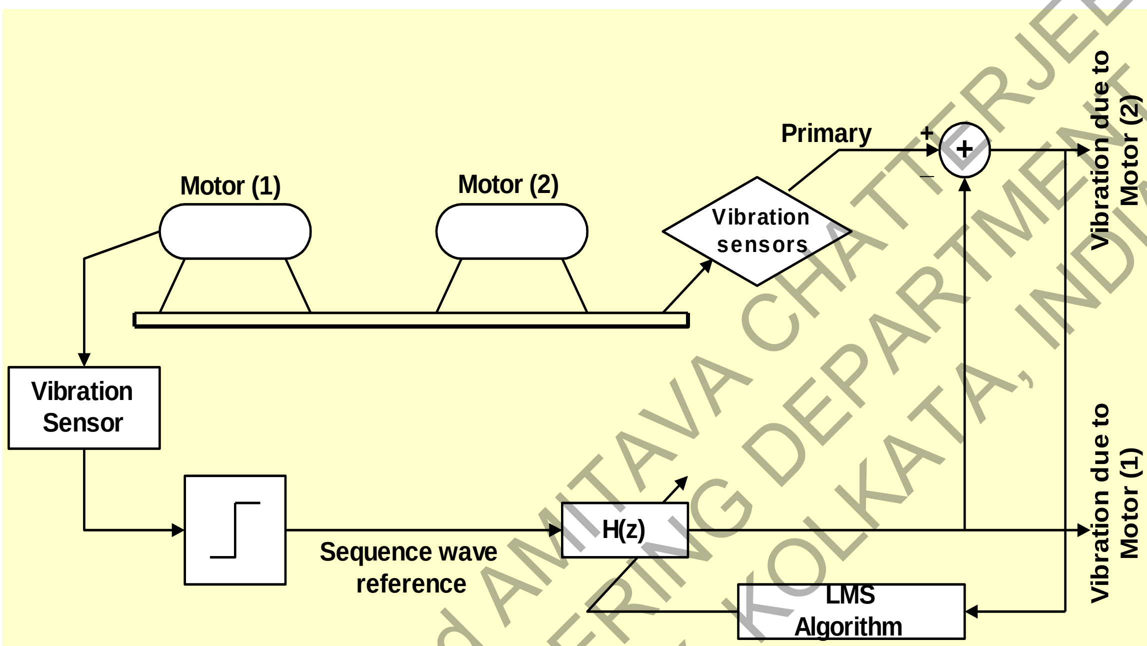


b) Adaptive noise cancellation in speech signals:

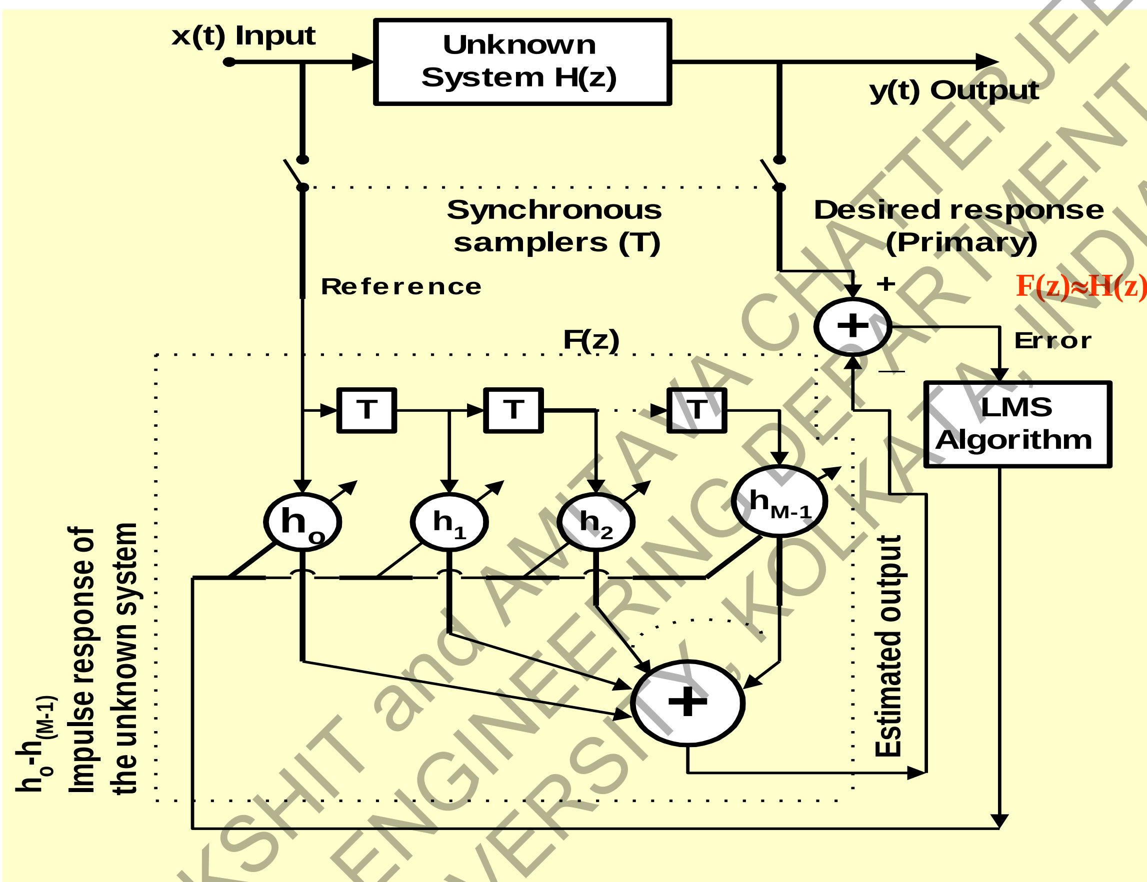


c) Vibration analysis (extraction of harmonic signals)

Separating vibrations from two variable speed motors:



d) FIR Modeling of an unknown system:



Here the unknown system H(z) is modeled by an FIR filter with M adjustable coefficients (h_0, h_1, … , h_{M-1}). The reference input to the FIR filter is identical to the system input or plant input i.e. x_n.

Here, estimated output of the filter $\hat{y}_n = \sum_{k=0}^{M-1} h_k x_{n-k} \quad (n = 0,1,2,\cdots)$

Hence, the error sequence $e_n = y_n - \hat{y}_n$

The coefficients h_k are selected such that $\xi_M = \sum_{n=0}^{N}\left[y_n - \sum_{k=0}^{M-1} h_k x_{n-k}\right]^2$ is minimum where (N+1) is the number of observations. This is called the least-squares criterion. This condition leads us to a set of linear equations, from which the filter coefficients can be determined, given as,

$$\sum_{k=0}^{M-1} h_k r_{xx}(j-k) = r_{yx}(j), \quad j = 0,1,\cdots,M-1$$

$r_{xx}(j)$: auto-correlation of the sequence x_n

$r_{yx}(j)$: cross-correlation of the system output y_n with the input sequence x_n

REFERENCES

[1] J.G. Proakis and D.G. Manolakis. Digital Signal Processing: Principles, Algorithms, and Applications. Fourth Edition, Pearson Education, 2007.

[2] R.J. Schilling and S.L. Harris. Fundamentals of Digital Signal Processing using MATLAB. Thomson India Edition, 2007.