# Adaptive Filtering: Part I

Dr. Deepa Kundur

University of Toronto





## Discrete-Time Signals and Systems

**Reference:**

Sections 13.1 and 13.2 of

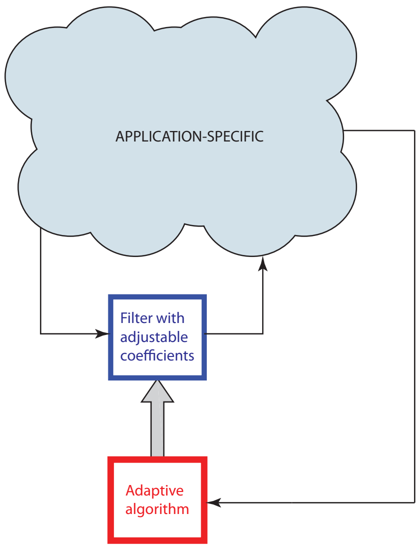
John G. Proakis and Dimitris G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th edition, 2007.





## Adaptive Filters

- filters with adjustable coefficients
- incorporate algorithms that allow the filter coefficients to adapt to signal statistics








## Adaptive Filters

- used when statistical characteristics of the signal to be filtered are either unknown a priori or are slowly time-variant
- two main aspects:
  - form of filter (FIR vs. IIR / direct-form vs. lattice-form); determines what the filter coefficients represent.
  - criterion for optimizing the adjustable filter parameters; determines how the filter coefficients are adapted.
- criterion
  - must provide meaningful measure of filter performance
  - must result in a practically realizable algorithm

# System Identification or System Modeling

- Goal: identify an unknown (possibly time-varying) system called a plant
- Model: the plant can be represented as a FIR filter with $M$ adjustable coefficients; see ▸ Figure 13.1.1 of text.





- The following signals are involved in the filter update process:
  - $x(n)$: input to the plant and FIR filter
  - $y(n)$: output from the plant
  - $\hat{y}(n)$: output from the FIR filter

$$\hat{y}(n) = \sum_{k=0}^{M-1} h(k)x(n-k)$$

  - $e(n) = y(n) - \hat{y}(n)$: error sequence used for filter coefficient update

  See ▸ Figure 13.1.2 of text.





- Assume our signals are observed for $n = 0, 1, 2, \dots, N$.
- Consider the following minimization criterion:

$$\begin{aligned}\mathcal{E}_M &= \sum_{n=0}^{N} e^2(n) = \sum_{n=0}^{N} [y(n) - \hat{y}(n)]^2 \\ &= \sum_{n=0}^{N} \left[ y(n) - \sum_{k=0}^{M-1} h(k)x(n-k) \right]^2\end{aligned}$$





- Least squares criterion (set of linear equations for determining filter coefficients):

$$\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) = r_{yx}(l)$$

  for $l = 0, 1, 2, \dots, M-1$.

$$\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) = r_{yx}(l) \quad \text{for } l = 0, 1, 2, \ldots, M-1$$

- $M$ unknowns: $h(0), h(1), h(2), \ldots, h(M-1)$
- $M$ equations:

$$\begin{aligned}
\sum_{k=0}^{M-1} h(k) r_{xx}(0-k) &= r_{yx}(0) & \text{Eq. (1)}\\
\sum_{k=0}^{M-1} h(k) r_{xx}(1-k) &= r_{yx}(1) & \text{Eq. (2)}\\
&\vdots\\
\sum_{k=0}^{M-1} h(k) r_{xx}(M-1-k) &= r_{yx}(M-1) & \text{Eq. } (M)
\end{aligned}$$





# Time-Average Autocorrelation/Cross-correlation

- $r_{xx}(n)$: time-average autocorrelation sequence of $x(n)$
- $r_{yx}(n)$: time-average crosscorrelation sequence between $y(n)$ and $x(n)$





$$\begin{aligned}
\mathcal{E}_M &= \sum_{n=0}^{N} \left[ y(n) - \sum_{k=0}^{M-1} h(k) x(n-k) \right]^2\\
\frac{\partial \mathcal{E}_M}{\partial h(l)} &= \sum_{n=0}^{N} 2 \left[ y(n) - \sum_{k=0}^{M-1} h(k) x(n-k) \right] \cdot (-1) \cdot x(n-l)\\
&= -2 \sum_{n=0}^{N} \left[ y(n) x(n-l) - \sum_{k=0}^{M-1} h(k) x(n-k) x(n-l) \right]\\
&= 2 \sum_{k=0}^{M-1} h(k) \underbrace{\sum_{n} x(n-k) x(n-l)}_{\equiv N\ r_{xx}(l-k)} - 2 \underbrace{\sum_{n} y(n) x(n-l)}_{\equiv N\ r_{yx}(l)}
\end{aligned}$$





ASIDE:

For an observation length of $N$ samples, the time-average crosscorrelation sequence between $a(n)$ and $b(n)$ is:

$$\begin{aligned}
r_{ab}(l) &\equiv \frac{1}{N} \sum_{n} a(n) b(n-l)\\
r_{ab}(l-k) &= \frac{1}{N} \sum_{n} a(n) b(n-(l-k)) \quad \text{let } n' = n+k\\
&= \frac{1}{N} \sum_{n'} a(n'-k) b(n'-l)
\end{aligned}$$

$$\frac{\partial \mathcal{E}_M}{\partial h(l)} = 2\sum_{k=0}^{M-1} h(k) \underbrace{\sum_n x(n-k)x(n-l)}_{\equiv N\ r_{xx}(l-k)} - 2\underbrace{\sum_n y(n)x(n-l)}_{\equiv N\ r_{yx}(l)}$$

$$= 2N\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) - 2N r_{yx}(l)$$

To determine extrema, set $\frac{\partial \mathcal{E}_M}{\partial h(l)} = 0$,

$$\cancel{2}\ \cancel{N}\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) - \cancel{2}\ \cancel{N} r_{yx}(l) = 0$$

$$\boxed{\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) = r_{yx}(l)}$$





**Q**: Is the resulting extrema, a minimum or maximum?
**A**: Let's look at the second derivative:

$$\frac{\partial \mathcal{E}_M}{\partial h(l)} = 2\sum_{k=0}^{M-1} h(k) \sum_n x(n-k)x(n-l) - 2\sum_n y(n)x(n-l)$$

$$= 2\sum_{k\neq l} h(k) \sum_n x(n-k)x(n-l) + 2h(l)\sum_n x(n-l)x(n-l) - 2\sum_n y(n)x(n-l)$$

$$\frac{\partial^2 \mathcal{E}_M}{\partial h^2(l)} = 0 + 2\sum_n x^2(n-l) - 0 = 2\sum_n x^2(n-l) > 0$$

MINIMA DETERMINED!





# System Identification or System Modeling

Therefore, the solution to:

$$\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) = r_{yx}(l)$$

minimizes a least squares criterion.

- If the plant is time-varying, then the FIR filter must continue to adapt such that is continues to model the time-varying system.
- This adaptation is governed by an algorithm ...

See Figure 13.1.2 of text.





# The LMS Algorithm: Background

- LMS = Least Mean Squares
- There is a common framework in all adaptive filtering applications. The least squares criterion leads to:

$$\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) = r_{dx}(l+D)$$

for $l = 0, 1, 2, \ldots, M-1$.

- Depending on the application $D$ may or may not be zero.

# The LMS Algorithm: Background

- Let us model $x(n)$ (input) and $d(n)$ (desired response) as random sequences.
- Assuming $x(n)$ and $d(n)$ are stationary and ergodic (time avarage = statistical average), $r_{xx}(n)$ and $r_{dx}(n)$ represent estimates of the:
  - true statistical autocorrelation, $\gamma_{xx}(n) \approx r_{xx}(n)$
  - true statistical crosscorrelation, $\gamma_{dx}(n) \approx r_{dx}(n)$
- We consider the true FIR filter coefficients to fulfill:

$$\sum_{k=0}^{M-1} \tilde{h}(k)\gamma_{xx}(l-k) = \gamma_{dx}(l+D)$$

for $l = 0, 1, 2, \ldots, M-1$.





# The LMS Algorithm: Background

$$\boxed{\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) = r_{dx}(l)}$$

- The coefficients $h(n)$ represent estimates of the true coefficients $\tilde{h}(n)$.
- Two challenges:
  1. Quality of the FIR coefficient estimate depends on the length of the data record $N$ available.
     - The estimate is statistically noisy.
     - Larger $N$ better.
  2. The underlying random sequence $x(n)$ is usually nonstationary, so the statistical correlations may vary with time.
     - The estimate is chasing a moving target.
     - Smaller $N$ better.





- Let us assume $x(n)$ is possibly complex-valued and consists of samples from a stationary random process with autocorrelation and crosscorrelation with $d(n)$:

$$\gamma_{xx}(m) = E[x(n)x^*(n-m)], \quad \gamma_{dx}(m) = E[d(n)x^*(n-m)]$$

- Analogous to our previous analysis,

$$\begin{aligned}
\hat{d}(n) &= \sum_{k=0}^{M-1} h(k)x(n-k) \\
e(n) &= d(n) - \hat{d}(n) = d(n) - \sum_{k=0}^{M-1} h(k)x(n-k) \\
\mathcal{E}_M &= E[|e(n)|^2] \qquad \text{MEAN -SQUARE ERROR}
\end{aligned}$$





$$\begin{aligned}
\mathcal{E}_M &= E[|e(n)|^2] \\
&= E\left[\left|d(n) - \sum_{k=0}^{M-1} h(k)x(n-k)\right|^2\right] \\
&= E\left[\left(d(n) - \sum_{k=0}^{M-1} h(k)x(n-k)\right)\left(d(n) - \sum_{l=0}^{M-1} h(l)x(n-l)\right)^*\right]
\end{aligned}$$

ASIDE:

The magnitude squared for the difference of complex numbers can be represented as follows for $a, b, c \in \mathbb{C}$ where $c = a - b$:

$$
\begin{aligned}
|c|^2 &= c \cdot c^* \\
&= (a-b)\cdot(a-b)^* = (a-b)\cdot(a^* - b^*) \\
&= a \cdot a^* - (a^* \cdot b + a \cdot b^*) + b \cdot b^* \\
&= |a|^2 - 2\mathrm{Re}\{a \cdot b^*\} + |b|^2
\end{aligned}
$$

Note:

$$
a^* \cdot b + a \cdot b^* = \underbrace{a^* \cdot b}_{\equiv d} + (a^* \cdot b)^* = d + d^* = 2\mathrm{Re}\{d\} = 2\mathrm{Re}\{a \cdot b^*\}
$$





$$
\begin{aligned}
\mathcal{E}_M &= E[|e(n)|^2] \\
&= E\left[\left| d(n) - \sum_{k=0}^{M-1} h(k)x(n-k) \right|^2\right] \\
&= E\left[\left( d(n) - \sum_{k=0}^{M-1} h(k)x(n-k) \right)\left( d(n) - \sum_{l=0}^{M-1} h(l)x(n-l) \right)^*\right] \\
&= E\left[ |d(n)|^2 - 2\mathrm{Re}\left\{ d(n) \left( \sum_{l=0}^{M-1} h(l)x(n-l) \right)^* \right\} \right. \\
&\qquad \left. + \sum_{k=0}^{M-1}\sum_{l=0}^{M-1} h^*(l)h(k)x^*(n-l)x(n-k) \right] \\
&= \underbrace{E[|d(n)|^2]}_{=\gamma_{dd}(0)} - 2\mathrm{Re}\left\{ \sum_{l=0}^{M-1} h^*(l) \underbrace{E[d(n)x^*(n-l)]}_{=\gamma_{dx}(l)} \right\} \\
&\qquad + \sum_{k=0}^{M-1}\sum_{l=0}^{M-1} h^*(l)h(k) \underbrace{E[x^*(n-l)x(n-k)]}_{=\gamma_{xx}(l-k)}
\end{aligned}
$$





$$
\begin{aligned}
\mathcal{E}_M &= \gamma_{dd}(0) - 2\mathrm{Re}\left\{ \sum_{l=0}^{M-1} h^*(l)\gamma_{dx}(l) \right\} + \sum_{k=0}^{M-1}\sum_{l=0}^{M-1} h^*(l)h(k)\gamma_{xx}(l-k) \\
&= \sigma_d^2 - 2\mathrm{Re}\left\{ \sum_{l=0}^{M-1} h^*(l)\gamma_{dx}(l) \right\} + \sum_{k=0}^{M-1}\sum_{l=0}^{M-1} h^*(l)h(k)\gamma_{xx}(l-k)
\end{aligned}
$$

Recall for the real and deterministic case, we considered:

$$
\begin{aligned}
\mathcal{E}_M &= \sum_{n=0}^{N}\left[ d(n) - \sum_{k=0}^{M-1} h(k)x(n-k) \right]^2 \\
&= N r_{dd}(0) - 2N \sum_{l=0}^{M-1} h(l) r_{dx}(l) + N \sum_{k=0}^{M-1}\sum_{l=0}^{M-1} h(l)h(k) r_{xx}(l-k)
\end{aligned}
$$





Recall, the solution for the real and deterministic case is:

$$
\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) = r_{dx}(l)
$$

for $l = 0, 1, \ldots, M-1$.

Similarly, for the more general case, we have:

$$
\sum_{k=0}^{M-1} h(k)\gamma_{xx}(l-k) = \gamma_{dx}(l)
$$

for $l = 0, 1, \ldots, M-1$.

- Latter equation is called the Wiener-Hopf equation.
- The filter coefficients $h(k)$ that solve the Wiener-Hopf equation represent the $\boxed{\text{Wiener filter}}$.

$$\sum_{k=0}^{M-1} h(k) r_{xx}(l-k) = r_{dx}(l) \quad (1)$$

$$\sum_{k=0}^{M-1} h(k) \gamma_{xx}(l-k) = \gamma_{dx}(l) \quad (2)$$

- Equation (2) makes use of the actual statistical autocorrelation and crosscorrelation to determine the filter coefficients.
  - yield optimum (Wiener) filter coefficients in the MSE sense
- Equation (1) makes use of estimates for the autocorrelation and crosscorrelation to determine the filter coefficients.
  - yield estimates of optimum filter coefficients ■