ADAPTIVE FILTERING

OUTLINE.

* APPLICATIONS OF ADAPTIVE FILTERS

* ADAPTIVE DIRECT FORM F.I.R. FILTERS
  (LMS, RLS algorithms)

* ADAPTIVE LATTICE - LADDER FILTERS

* PERFORMANCE COMPARISON

ADAPTIVE FILTERS

The statistical characteristics of the signals to be filtered are either unknown a priori or, slowly time variant (non-stationary signals)

* Adaptive beamforming [Widrow et al. (1967)]
* Adaptive equalization [Lucky (1965), Proakis et al. (1969, 1970, 1975), Gersho (1969), Picinbono (1977)]
* Adaptive noise canceling [Widrow et al. (1975), Hsu and Giordano (1978), Ketsum and Proakis (1982)]
* System modeling / identification, Echo cancellation Speech coding, etc.

REFERENCES

A few key references from the extensive literature on adaptive filtering are:

* SIMON HAYKIN, "Adaptive Filter Theory", Prentice-Hall (1991)

* J. PROAKIS et al., "Advanced Digital Signal Processing" (Macmillan, 1992)

* S. QURESHI, "Adaptive Equalization", Proc. of the IEEE, Vol. 73, No. 9, Sept. 1985.

FIR FILTER for adaptive filtering

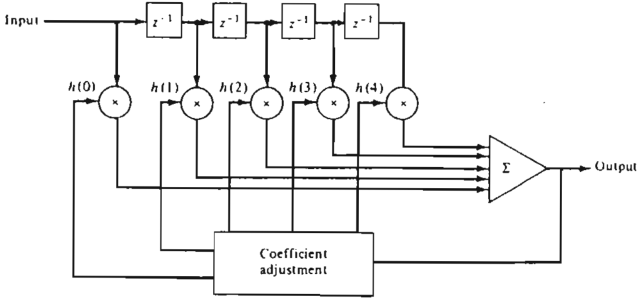


FIGURE 1 Direct-form adaptive FIR filter.

* Stability of the filter depends on coefficient adjustment algorithm
* IIR filters suffer from stability problems more often.
* Direct form and lattice form FIR filter structures are common

OPTIMIZATION CRITERIA

* Very important for efficient adjustment of filter coefficients

* Criterion must be a meaningful measure of filter performance and result in practically realizable algorithms

* The least-squares (LS) and mean-square error (MSE) criteria result in quadratic performance index with a single minimum. They are used widely in practice.

SYSTEM IDENTIFICATION / MODELING

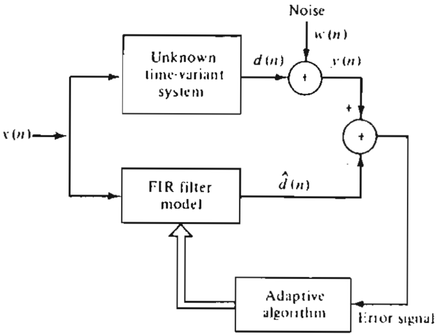


FIGURE 2 Application of adaptive filtering to system identification.

* $\hat{d}(n) = \sum_{k=0}^{M-1} h(k)\, x(n-k)$
* $e(n) = y(n) - \hat{d}(n)$ is an error sequence
* Select $\{h(k)\}_{k=0,\dots,M-1}$ to minimize $\sum_{n=0}^{N-1} |e(n)|^2$ (LS criterion)

ADAPTIVE CHANNEL EQUALIZATION (1)

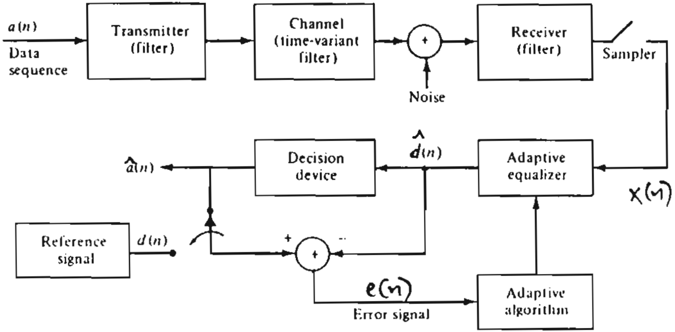


FIGURE 3 Application of adaptive filtering to adaptive channel equalization.

$$* \quad x(n) = \sum_{k=0}^{\infty} a(k)\cdot g(n-k) + w(n) = a(n) + \underbrace{\sum_{\substack{k=0 \\ k\neq n}}^{\infty} a(k)\, g(n-k)}_{\text{Intersymbol Interference (ISI)}} + w(n)$$

true symbol

noise.

ADAPTIVE CHANNEL EQUALIZATION (2)

- Assume that the equalizer filter is an FIR filter with M adjustable coefficients $\{h(n)\}_{n=0,\ldots,M-1}$

- Output of equalizer: $\hat{d}(n) = \sum_{k=0}^{M-1} h(k)\, x(n-k)$

- Form the error: $e(n) = d(n) - \hat{d}(n)$

  (where: $d(n) = a(n+D)$ is the desired true value and D accounts for delay in the channel)

  (Note: $d(n) = \hat{a}(n)$ after initial convergence is achieved)

- Select $\{h(n)\}_{n=0,\ldots,M-1}$ to minimize $\sum_{n=0}^{N-1} |e(n)|^2$

ECHO CANCELLATION (1)

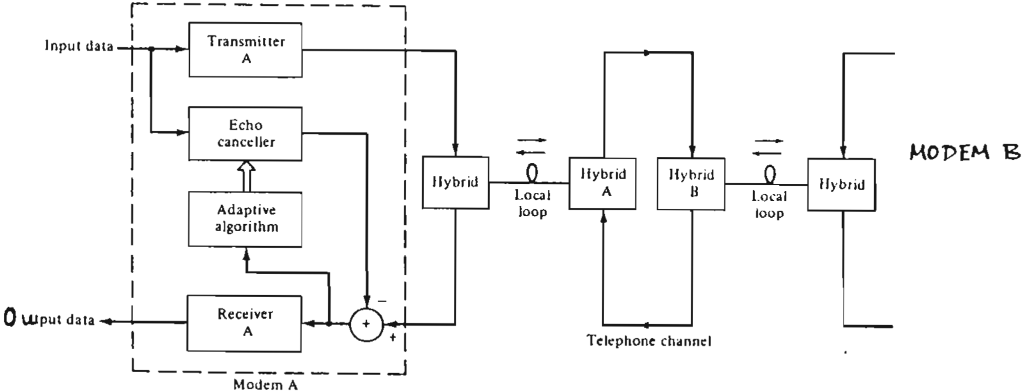
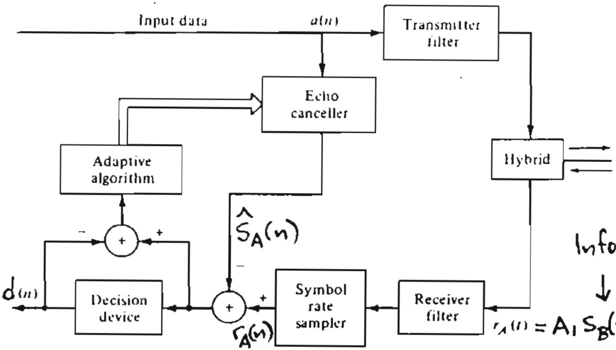


FIGURE 3. Block diagram model of a digital communication system that used echo cancellers in the modems

* Echoes due to impedance mismatch between Hybrid A and the channel (near-end echoes)
* Echoes due to impendanc mismatch at Hybrid B (far-end echoes)

ECHO CANCELLATION (2)

MODEM A



Near and far end echoes

Information

$$r_A(t) = A_1 S_B(t) + A_2 S_A(t-d_1) + A_3 S_A(t-d_2)$$

$$+ w(t)$$

noise.

FIGURE 4. Symbol-rate echo canceller.

* $\hat{S}_A(n) = \sum_{k=0}^{M-1} h(k)\, a(n-k)$,

* Error: $e(n) = d(n) - [r_A(n) - \hat{S}_A(n)]$, Minimize $\sum_{n=0}^{N-1} |e(n)|^2$

SUPPRESSION OF NARROWBAND INTERFERENCE IN A WIDEBAND SIGNAL (1)

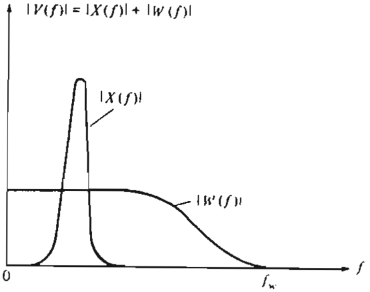


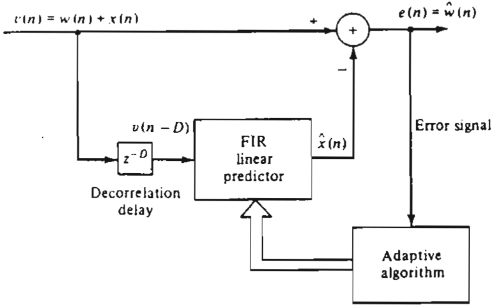
FIGURE 5. Strong narrowband interference $X(f)$ in a wideband signal $W(f)$.

$*$ $v(n) = x(n) + w(n)$ ← not highly correlated.

Highly correlated

($x(n)$ and $w(n)$ uncorrelated)

SUPPRESSION OF NARROWBAND INTERFERENCE IN A WIDEBAND SIGNAL (2)



* $\hat{x}(n) = \sum_{k=0}^{M-1} h(k)\, v(n-D-k)$

* Form the error

$$e(n) = v(n) - \hat{x}(n)$$

* Obtain $\{h(k)\}_{k=0,1,\ldots,M-1}$ by minimizing the LS criterion

$$\sum_{n=0}^{N-1} |e(n)|^2$$

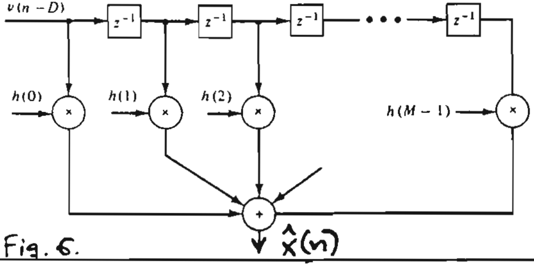


Fig. 6.

ADAPTIVE DIRECT-FORM FIR FILTERS

From previous examples we observe a common framework in adaptive filter applications.

→ Given,

- the observverd (received) data samples $\{x(n)\}$,
- an FIR digital filter with unknown coefficients $\{h(n)\}_{n=0,\ldots,M-1}$
- A desired response for the filter $\{d(n)\}$

→ Form the error quantity: $e(n) = d(n) - \sum_{k=0}^{M-1} h(k)\, x(n-k)$

→ Minimize a function of $e(n)$ with respect to the $\{h(k)\}$

(Minimization should be carried out adaptively to accomodate changing signal conditions).

MINIMUM MEAN SQUARE ERROR CRITERION (MMSE)

Minimize the MSE function:

$$\boxed{J(\underline{h}_M) = E\{|e(n)|^2\}}$$

where, $\underline{h}_M = [h(0), \ldots, h(M-1)]^T$,

$$e(n) = d(n) - \sum_{k=0}^{M-1} h(k)\, x(n-k)$$

SOLUTION: ($J(\cdot)$ is a quadratic function of $\underline{h}_M$)

$$\left[\sum_{k=0}^{M-1} h(k) R_{xx}(\ell - k) = r_{dx}(\ell) \quad , \quad \ell = 0, 1, \ldots, M-1 \right]$$

where: $R_{xx}(m) = E\{x(n)x^*(n-m)\}$, $r_{dx}(m) = E\{d(n)x^*(n-m)\}$

MMSE criterion (2)

In matrix form the solution is written as:

$$\underline{R}_M \cdot \underline{h}_M = \underline{r}_d \qquad \text{(Wiener - Hopf Equation)}$$

where:

$$\underline{R}_M = \begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & \cdots & R_{xx}(-M+1) \\ R_{xx}(1) & R_{xx}(0) & \cdots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(M-1) & \cdot & \cdots & R_{xx}(0) \end{bmatrix}, \quad \underline{h}_M = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(M-1) \end{bmatrix}, \quad \underline{r}_d = \begin{bmatrix} r_{dx}(0) \\ r_{dx}(1) \\ \vdots \\ r_{dx}(M-1) \end{bmatrix}$$

Autocorrelation Matrix (MxM), Filter coefficients (Mx1), Crosscorrelation Vector (Mx1)

MMSE Criterion (3)

* Optimum solution: $\underline{h}_{opt} = \underline{R}_M^{-1} \cdot \underline{r}_d$

* Minimum mean square error:

$$J_{min} = J(\underline{h}_{opt}) = E\{|d(n)|^2\} - \sum_{k=0}^{M-1} h_{opt}(k)\, r_{dx}^*(k)$$

$$= \sigma_d^2 - \underline{r}_d^H \underline{R}_M^{-1} \underline{r}_d$$

[* denotes conjugation
H denotes conjugate transpose]

* $\underline{R}_M$ is Hermitian and Toeplitz: Efficient solutions exist.

LEAST SQUARES CRITERION (LS)

Minimize the least-squares function

$$\boxed{J_{LS}(\underline{h}_M) = \sum_{n=0}^{N-1} |e(n)|^2}$$

where, $\underline{h}_M$ and $e(n)$ are as before.

SOLUTION: ($J_{LS}(\cdot)$ is also a quadratic function of $\underline{h}_M$)

$$\left[\sum_{k=0}^{M-1} h(k) \cdot \hat{R}_{xx}(\ell-k) = \hat{r}_{dx}(\ell+D) \quad , \quad \ell = 0, 1, \ldots, M-1 \right]$$

where: $\hat{R}_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) x^*(n-m)$, $\hat{r}_{dx}(m) = \frac{1}{N} \sum_{n=0}^{N-1} d(n) x^*(n-m)$

MMSE and LS CRITERIA

* The solutions obtained from both criteria are similar in form
* In MMSE, the true statistical autocorrelation and crosscorrelation are employed. The optimum (Wiener) filter coefficients are obtained.
* In LS, estimates of the autocorrelation and cross-correlation are used. The underlying assumption is that the observed data sequence is stationary and ergodic. Estimates of the optimum filter coefficients are obtained.

RECURSIVE METHODS BASED ON THE MMSE

* Consider the recursive algorithm

$$\underline{h}_M(n+1) = \underline{h}_M(n) + \frac{1}{2}\mu(n)\underline{D}(n) \quad , \quad n=0,1,\ldots$$

where: $\underline{h}_M(n)$ the vector of filter coefficients at iteration n

$\mu(n)$ is a step size at iteration n

$\underline{D}(n)$ iteration vector at iteration n.

* <u>SPECIAL CASE</u>: Steepest-descent methods (gradient methods)

$$\underline{D}(n) = -\frac{dJ(\underline{h}_M(n))}{d\underline{h}_M(n)} = 2\left[\underline{r}_d - \underline{R}_M \cdot \underline{h}_M(n)\right].$$

Thus:

$$\boxed{\underline{h}_M(n+1) = \underline{h}_M(n) + \mu(n)\left[\underline{r}_d - \underline{R}_M \underline{h}_M(n)\right]} \quad , n=0,1,\ldots$$

RECURSIVE METHODS (MMSE) (2)

By substituting, $\underline{r}_d = E\{d(n)\cdot \underline{X}_M^*(n)\}$, $\underline{R}_M = E\{\underline{X}_M^*(n)\cdot \underline{X}_M^T(n)\}$

where $\underline{X}_M(n) = [x(n), x(n-1), \dots, x(n-M+1)]^T$

we obtain:

$$\underline{h}_M(n+1) = \underline{h}_M(n) + \mu(n)\cdot E\Big\{\underline{X}_M^*(n)\big[\underbrace{d(n) - \underline{X}_M^T(n)\,\underline{h}_M(n)}_{e(n)}\big]\Big\}$$

Thus:

$$\boxed{\underline{h}_M(n+1) = \underline{h}_M(n) + \mu(n)\cdot E\{e(n)\underline{X}_M^*(n)\}}\quad, n=0,1,\dots$$

* It can be shown that the above algorithm converges provided $\mu(n)$ is properly chosen
* Adaptation stops when $E\{e(n)\underline{X}_M^*(n)\}=0$ (orthogonality principle)

THE LEAST-MEAN-SQUARES (LMS) ALGORITHM

* In practice the term $E\{e(n)\underline{X}_M^*(n)\}$ is replaced by an estimate.
* A simple unbiased estimate is obtained by dropping the expectation operation
* In practice, the step size $\mu(n)$ is fixed to a constant value $\mu > 0$.

Thus,

LMS: $$\boxed{\underline{h}_M(n+1) = \underline{h}_M(n) + \mu \cdot e(n) \cdot \underline{X}_M^*(n)}, \quad n = 0, 1, \ldots$$

(Various variations of the LMS algorithm exist in the literature).

PROPERTIES OF THE LMS ALGORITHM (1)

- The rate of convergence depends on the following:
  1) Step size μ: The higher the value of μ the faster the convergence. The higher the value of μ the higher the final mean square error achieved by the algorithm.
  2) Eigenvalue spread of R_M: The larger the eigenvalue spread the slower the convergence.
- There is a trade-off between convergence speed and final mean square error.
- The algorithm is stable provided

$$0 < \mu < \frac{2}{\lambda_{max}}$$

where λ_{max} is the largest eigenvalue of R_M

PROPERTIES OF THE LMS ALGORITHM (2)

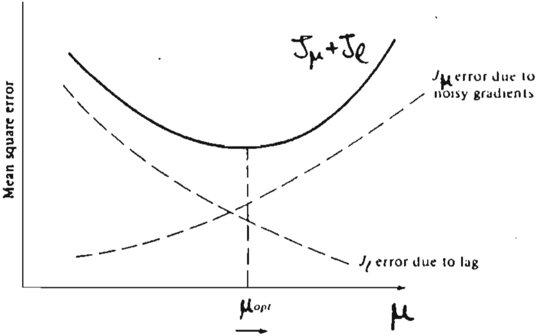
- In practice choose: $0 < \mu < \dfrac{1}{\left(\underline{X}_M^T \underline{X}_M^*\right)} = \dfrac{1}{\sum_{k=0}^{M-1} |x(n-k)|^2}$

- In nonstationary signal environments (slowly time varying) the final mean-square error achieved is

$$J_{total}(n) = J_{min}(n) + J_{\mu}(n) + J_{\ell}(n)$$

$J_{\mu}(n)$: gradient noise error

$J_{\ell}(n)$: lag error



LMS ALGORITHM: EXAMPLE (Channel equalization)

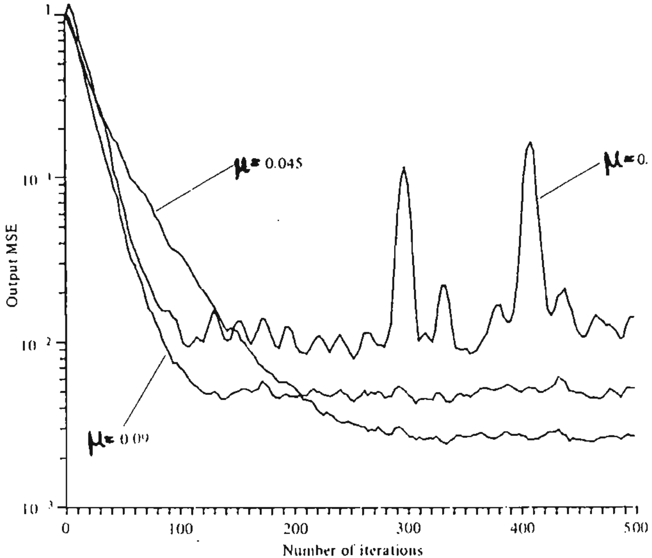


FIGURE 9. Learning curves for the LMS algorithm applied to an adaptive equalizer of length $M = 11$ and a channel with eigenvalue spread $\lambda_{max}/\lambda_{min} = 11$.

SUMMARY OF THE LMS ALGORITHM.

Parameters: M = number of taps
μ = step size $\left(0 < \mu < \dfrac{2}{\sum_{i=0}^{M-1} |x(n-i)|^2}\right)$

Initial Conditions: $\underline{h}_M(0) = \underline{0} = [0, 0, \ldots, 0]^T$

Data: $\underline{X}_M(n) = [x(n), x(n-1) \ldots x(n-M+1)]^T$

$d(n)$: desired response.

$\underline{h}_M(n) = [h(0,n), h(1,n), \ldots, h(M-1,n)]^T$

Computation: For $n = 0, 1, 2, \ldots$ compute

$$\left[\begin{array}{l} e(n) = d(n) - \underline{X}_M^T(n)\, \underline{h}_M(n) \\[2ex] \underline{h}_M(n+1) = \underline{h}_M(n) + \mu \cdot e(n) \cdot \underline{X}_M^*(n) \end{array}\right]$$

RECURSIVE LEAST SQUARES (RLS) ESTIMATION

Given an FIR filter with coefficients

$$\underline{h}_M(n) = [h(0,n), h(1,n), \ldots, h(M-1,n)]^T$$

and the data vector

$$\underline{X}_M(n) = [x(n), x(n-1), \ldots, x(n-M+1)]^T$$

Suppose we observe the vectors: $\underline{X}_M(l)$, $l = 0, 1, 2, \ldots, n$ and we wish to determine the filter coefficients vector $\underline{h}_M(n)$ that minimizes the weighted sum of magnitude-squared errors:

$$\left[\mathcal{E}_M = \sum_{l=0}^{n} w^{n-l} |e_M(l,n)|^2 \right], \quad 0 \leq w \leq 1$$

where, $e_M(l,n) = d(l) - \hat{d}(l)$ and w is a forgetting factor.

RLS ESTIMATION (2)

Minimization of $\mathcal{E}_M$ with respect to the $\underline{h}_M(n)$ yields

$$\underline{R}_M(n)\cdot\underline{h}_M(n)=\underline{D}_M(n)$$

where,

$$\underline{R}_M(n)=\sum_{\ell=0}^{n} w^{n-\ell}\,\underline{X}_M^*(\ell)\,\underline{X}_M^T(\ell)$$

$$\underline{D}_M(n)=\sum_{\ell=0}^{n} w^{n-\ell}\,\underline{X}_M^*(n)\cdot d(\ell)$$

SOLUTION:

$$\left[\ \underline{h}_M(n)=\underline{R}_M^{-1}\,\underline{D}_M(n)\ \right]$$

RLS ESTIMATION (3)

Suppose we have the optimum solution at time $n-1$ and we wish to compute $\underline{h}_M(n) \Longrightarrow$ Recursive solution

* TIME UPDATE EQUATIONS

$$\underline{R}_M(n) = w\,\underline{R}_M(n-1) + \underline{X}_M^*(n)\,\underline{X}_M^T(n)$$

$$\underline{D}_M(n) = w\,\underline{D}_M(n-1) + \underline{X}_M^*(n) \cdot d(n)$$

* MATRIX INVERSION LEMMA

$$\underline{R}_M^{-1}(n) = \frac{1}{w}\left[\underline{R}_M^{-1}(n-1) - \frac{\underline{R}_M^{-1}(n-1)\cdot\underline{X}_M^*(n)\,\underline{X}_M^T(n)\,\underline{R}_M^{-1}(n-1)}{w + \underline{X}_M^T(n)\,\underline{R}_M^{-1}(n-1)\,\underline{X}_M^*(n)}\right]$$

RLS ESTIMATION (4)

Let $\underline{P}_M(n) = \underline{R}_M^{-1}(n)$, and $\underline{K}_M(n) = \dfrac{\underline{P}_M(n-1)\,\underline{X}_M^*(n)}{w + \underline{X}_M^T(n)\,\underline{P}_M(n-1)\,\underline{X}_M^*(n)}$

(Kalman Gain vector)

Then,

$$\underline{h}_M(n) = \underline{R}_M^{-1}(n)\cdot\underline{D}_M(n) = \underline{P}_M(n)\cdot\underline{D}_M(n) = \cdots$$

$$= \underbrace{\underline{P}_M(n-1)\,\underline{D}_M(n-1)}_{\underline{h}_M(n-1)} + \underline{K}_M(n)\underbrace{\left[d(n) - \underline{X}_M^*(n)\cdot\underline{h}_M(n)\right]}_{e_M(n)}$$

or

$$\boxed{\underline{h}_M(n) = \underline{h}_M(n-1) + \underline{K}_M(n)\,e_M(n)}$$

RLS ESTIMATION (5)

It can be shown that $\underline{K}_M(n) = \underline{P}_M(n)\underline{X}_M^*(n)$.

By substituting into RLS recursion equation.

$$\boxed{\underline{h}_M(n) = \underline{h}_M(n-1) + \underline{P}_M(n)\,\underline{X}_M^*(n)\,e_M(n)} \qquad \text{RLS}$$

Note that for the LMS algorithm we found

$$\boxed{\underline{h}_M(n) = h_M(n-1) + \mu \cdot \underline{X}_M^*(n)\,e_M(n)} \qquad \text{LMS}$$

RLS ALGORITHM

$[\, \underline{h}_M(0) = \underline{0} \,,\ \underline{P}_M(-1) = \frac{1}{\delta} I_M \,]$

1) Compute the filter output: $\hat{d}(n) = \underline{X}_M^T(n)\, \underline{h}_M(n-1)$

2) Compute the error : $e_M(n) = d(n) - \hat{d}(n)$

3) Compute the gain vector:

$$\underline{K}_M(n) = \frac{\underline{P}_M(n-1)\, \underline{X}_M^*(n)}{w + \underline{X}_M^T(n)\, \underline{P}_M(n)\, \underline{X}_M^*(n)}$$

4) Update the inverse of the autocorrelation matrix.

$$\left[\underline{P}_M(n-1) - \underline{K}_M(n)\, \underline{X}_M^*(n)\, \underline{P}_M(n-1)\right]$$

5) Update the coefficient vector of the filter.

$$h_M(n) = \underline{h}_M(n-1) + \underline{K}_M(n)\, e_M(n)$$

RLS vs LMS : Example

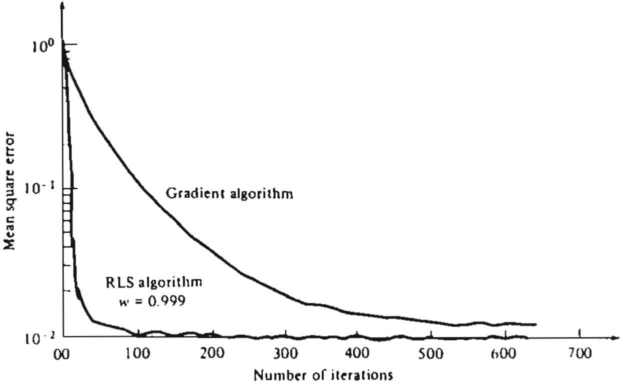


FIGURE 10 Learning curves for RLS algorithm and LMS algorithm for adaptive equalizer of length $M = 11$. The eigenvalue spread of the channel is $\lambda_{max}/\lambda_{min} = 11$. The step size for the LMS algorithm is $\Delta = 0.02$. (From *Digital Communication* by John G. Proakis. © 1983 by McGraw-Hill Book Company.

RLS algorithm

Limitations: → Complexity $\sim M^3$

→ Stability of the algorithm requires high precision arithmetic (24 bits or more) (Computation of $\underline{P}_M(n)$)

Potential Solution:

Use a square-root RLS algorithm based on LDU decomposition of $\underline{R}_M(n)$ of $\underline{P}_M(n)$