Convergence of Adaptive Algorithms

B Swaroop Reddy and Dr G V V Sharma*

Contents

***Abstract*—This manual provides theoretical insights into adaptive algorithms.**

1 Wiener-Hopf Equation

Problem 1.1. Let

$$e(n) = d(n) - W^T(n)X(n) \tag{1.1}$$

Show that

$$E[e^2(n)] = r_{dd} - W^T(n)r_{xd} - r_{xd}^T W(n) + W^T(n)RW(n) \tag{1.2}$$

where

$$r_{dd} = E[d^2(n)] \tag{1.3}$$
$$r_{xd} = E[X(n)d(n)] \tag{1.4}$$
$$R = E[X(n)X^T(n)] \tag{1.5}$$

Problem 1.2. By computing

$$\frac{\partial J(n)}{\partial W(n)} = 0, \tag{1.6}$$

show that the optimal solution for

$$W^*(n) = \min_{W(n)} E\left[e^2(n)\right] = R^{-1}r_{xd} \tag{1.7}$$

This is the Wiener optimal solution.

*The author is with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India e-mail: gadepall@iith.ac.in.

2 Convergence of the LMS Algorithm

2.1 Convergence in the Mean

Problem 2.1. Show that R in (1.5) is symmetric as well as positive definite.

Let

$$\tilde{W}(n) = W(n) - W_* \tag{2.1}$$

where W_* is obtained in (1.7). Also, according to the LMS algorithm,

$$W(n+1) = W(n) + \mu X(n)e(n) \tag{2.2}$$
$$e(n) = d(n) - X^T(n)W(n) \tag{2.3}$$

Problem 2.2. Show that

$$\tilde{E}\left[W(n+1)\right] = [I - \mu R]E\left[\tilde{W}(n)\right] \tag{2.4}$$

Problem 2.3. Show that

$$R = U\Lambda U^T \tag{2.5}$$

for some U, Λ, such that Λ is a diagonal matrix and $U^T U = I$.

Problem 2.4. Show that

$$\lim_{n\to\infty} \tilde{E}\left[W(n+1)\right] = 0 \iff \lim_{n\to\infty} [I - \mu\Lambda]^n = 0 \tag{2.6}$$

Problem 2.5. Using (2.6), show that

$$0 < \mu < \frac{2}{\lambda_{\max}} \tag{2.7}$$

where $\lambda_{\max}$ is the largest entry of Λ.

2.2 Convegenge in Mean-square sense

Problem 2.6. How can we choose the value of μ if LMS algorithm converges in mean-square sense.

Solution: $0 < \mu < \dfrac{1}{M(SignalPower)}$

Problem 2.7. Prove the result of the Problem 2.6.

Solution:

$$
\begin{aligned}
J(n) &= E[e^2(n)]\\
&= E[(d(n) - W^T(n)X(n))^2]\\
&= E[(d(n) - W^T(n)X(n))^T(d(n) - W^T(n)X(n))]\\
&= E[(d(n) - W_*^T X(n) - W^T(n)X(n) + W_*^T X(n))^T\\
&\quad (d(n) - W_*^T X(n) - W^T(n)X(n) + W_*^T X(n))]\\
&= E[(e_*(n) - \tilde{W}(n)X(n))^T(e_*(n) - \tilde{W}(n)X(n))]\\
&= E[e_*^2(n)] + E[\tilde{W}(n)X(n)X(n))^T\tilde{W}(n)] -\\
&\quad E[\tilde{W}(n)X(n)e_*(n)] - E[e_*(n)X^T(n)\tilde{W}(n)]
\end{aligned}
$$

Where

$$
\begin{aligned}
\tilde{W}(n) &= W(n) - W_*\\
e_*(n) &= d(n) - W_* X(n)
\end{aligned}
$$

$$E[e_*(n)X^T(n)(n)\tilde{W}(n)]$$

$$
\begin{aligned}
&= E[(d(n) - W_*^T X(n)X^T(n)(n)\tilde{W}(n))]\\
&= E[d(n)X^T(n)(n)\tilde{W}(n)] - E[W_*^T X(n)X^T(n)(n)\tilde{W}(n)]\\
&= [r_{xd} - W_*^T R]\tilde{W}(n) = 0
\end{aligned}
$$

$$
\begin{aligned}
J(n) &= E[e_*^2(n)] + E[\tilde{W}(n)X(n)X(n))^T\tilde{W}(n)]\\
&= J_{min}(n) + J_{ex}(n)
\end{aligned}
$$

Assumptions
$X(n)X^T(n) \perp\!\!\!\perp \tilde{W}(n)$

$$E[\tilde{W}^T(n)X(n)X^T(n)\tilde{W}(n)]$$

$$
\begin{aligned}
&= E[\sum_{i=1}^{M}\sum_{j=1}^{M}\tilde{W}_i(n)X_i(n)X_j(n)\tilde{W}_j(n)]\\
&= \sum_{i=1}^{M}\sum_{j=1}^{M}E[\tilde{W}_i(n)X_i(n)X_j(n)\tilde{W}_j(n)]\\
&= \sum_{i=1}^{M}\sum_{j=1}^{M}E[\tilde{W}_i(n)\tilde{W}_j(n)]E[X_i(n)X_j(n)]\\
&= E[\sum_{i=1}^{M}\sum_{j=1}^{M}\tilde{W}_i(n)E[X_i(n)X_j(n)]\tilde{W}_j(n)]\\
&= E[\tilde{W}^T(n)R\tilde{W}(n)]
\end{aligned}
$$

$$J(n) = J_{min} + E[\tilde{W}^T(n)R\tilde{W}(n)]$$

Since R is Symmetric and positive semidefinite, $R = U\Lambda U^T$, Where $U = [u_1, u_2, ..., u_M]$ and $\lambda_i > 0$ for i =1,2,3,...,M

$$
\begin{aligned}
J_{ex}(n) &= E[\tilde{W}^T(n)U)(U^T\tilde{W}(n))]\\
&= \sum_{i=1}^{M}\lambda_i E[\tilde{W}_u^i\tilde{W}_u^i]\\
&= \sum_{i=1}^{M}\lambda_i p_u^{ii}(n)
\end{aligned}
$$

$$\tilde{W}(n+1) = [I-\mu X(n)X^T(n)]\tilde{W}(n)+\mu X(n)[d(n)-X^T(n)W_*] \tag{2.8}$$

Let $P(n) = E[\tilde{W}(n)\tilde{W}^T(n)]$
and $P_u(n) = U^T P(n)U$
$p_u^{ii}(n)$ is the ii^{th} diagonal entry of $P_u(n)$
$p_u^{ii}(n+1) = [1 - 2\mu\lambda_i]p_u^{ii}(n) + \mu^2 J_{min}\lambda_i$

$$
\begin{aligned}
| 1 - 2\mu\lambda_i | &< 1\\
-1 < 1 - 2\mu\lambda_i &< 1\\
0 < \mu &< \frac{1}{\lambda_i}\\
0 < \mu < \frac{1}{\lambda_{max}} &< \frac{1}{\lambda_i}\\
0 < \mu &< \frac{1}{M(SignalPower)}
\end{aligned}
$$

Problem 2.8. Take a 2 X 1 matrix for $\tilde{W}(n)$ and X(n) show that $E[\tilde{W}^T(n)X(n)X^T(n)\tilde{W}(n)] = E[\tilde{W}^T(n)R\tilde{W}(n)]$

Solution: Let

$$X(n) = \begin{bmatrix} X_1(n)\\ X_2(n)\end{bmatrix}_{2X1} \quad W(n) = \begin{bmatrix} \tilde{W}_1(n)\\ \tilde{W}_2(n)\end{bmatrix}_{2X1} \tag{2.9}$$

$$LHS = E\left\{\tilde{W}^T(n)X(n)X^T(n)\tilde{W}(n)\right\}$$

$$
\begin{aligned}
&= E\left\{\sum_{i=1}^{2}\sum_{j=1}^{2}\tilde{W}_i(n)X_i(n)X_j(n)\tilde{W}_j(n)\right\}\\
&= E\left\{\tilde{W}_1^2(n)X_1^2(n) + 2\tilde{W}_1(n)\tilde{W}_2(n)X_1(n)X_2(n) + \tilde{W}_2^2(n)X_2^2(n)\right\}\\
&= E[\tilde{W}_1^2(n)]E[X_1^2(n)] + 2\tilde{E}[W_1(n)\tilde{W}_2(n)]E[X_1(n)X_2(n)]\\
&\quad + E[\tilde{W}_2^2(n)]E[X_2^2(n)]\\
&= E\left\{\tilde{W}_1^2(n)E[X_1^2(n)] + 2\tilde{W}_1(n)\tilde{W}_2(n)E[X_1(n)X_2(n)]\right.\\
&\quad \left. + \tilde{W}_2^2(n)E[X_2^2(n)]\right\}
\end{aligned}
$$

$$
\begin{aligned}
R &= E[X(n)X^T(n)] \\
&= E\left\{ \begin{bmatrix} X_1^2(n) & X_1(n)X_2(n) \\ X_1(n)X_2(n) & X_2^2(n) \end{bmatrix} \right\} \\
&= \begin{bmatrix} E[X_1^2(n)] & E[X_1(n)X_2(n)] \\ E[X_1(n)X_2(n)] & E[X_2^2(n)] \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
RHS &= E[\tilde{W}^T(n)R\tilde{W}(n)] \\
&= E\Big\{\tilde{W}_1^2(n)E[X_1^2(n)] + 2\tilde{W}_1(n)\tilde{W}_2(n) \\
&\quad E[X_1(n)X_2(n)] + \tilde{W}_2^2(n)E[X_2^2(n)]\Big\}
\end{aligned}
$$

Problem 2.9. Let $P(n) = E[\tilde{W}(n)\tilde{W}^T(n)]$ and $P_u(n) = U^T P(n)U$. Show that

$$P_u(n+1) = \Big(I - 2\mu\Lambda\Big)P_u(n) + \mu^2 J_{min}\Lambda$$

Solution: From equation 2.8

$$
\begin{aligned}
\tilde{W}(n+1) &= [I - \mu X(n)X^T(n)]\tilde{W}(n) \\
&\quad + \mu X(n)[d(n) - X^T(n)W_*] \\
&= [I - \mu X(n)X^T(n)]\tilde{W}(n) + \mu X(n)e_*(n)
\end{aligned}
$$

$$E\Big[\tilde{W}(n+1)\tilde{W}^n(n+1)\Big]$$

$$
\begin{aligned}
&= E\Big[(I - \mu X(n)X^T(n))\tilde{W}(n)\tilde{W}^T(n)(I - \mu X(n)X^T(n))^T \\
&\quad + \mu X(n)e_*(n)\tilde{W}(n)\tilde{W}^T(n)(I - \mu X(n)X^T(n))^T \\
&\quad + (I - \mu X(n)X^T(n))\tilde{W}(n)\mu e_*^T(n)X^T(n) \\
&\quad + \mu^2 X(n)e_*(n)e_*^T(n)X^T(n)\Big] \\
&= E\Big[(I - \mu X(n)X^T(n))\tilde{W}(n)\tilde{W}^T(n)(I - \mu X(n)X^T(n))^T\Big] \\
&\quad + E\Big[\mu X(n)e_*(n)\tilde{W}(n)\tilde{W}^T(n)(I - \mu X(n)X^T(n))^T\Big] \\
&\quad + E\Big[(I - \mu X(n)X^T(n))\tilde{W}(n)\mu e_*^T(n)X^T(n)\Big] \\
&\quad + E\Big[\mu^2 X(n)e_*(n)e_*^T(n)X^T(n)\Big]
\end{aligned}
$$

1^{st} term -

$$E\Big[(I - \mu X(n)X^T(n))\tilde{W}(n)\tilde{W}^T(n)(I - \mu X(n)X^T(n))^T\Big]$$

$$
\begin{aligned}
&= E\Big[\Big(\tilde{W}(n)\tilde{W}^T - \mu X(n)X^T(n)\tilde{W}(n)\tilde{W}^T\Big) \\
&\qquad \Big(I - \mu X(n)X^T(n)\Big)^T\Big] \\
&= E\Big[\tilde{W}(n)\tilde{W}^T(n) - \mu X(n)X^T(n)\tilde{W}(n)\tilde{W}^T(n) \\
&\qquad - \mu\tilde{W}(n)\tilde{W}^T(n)\mu X(n)X^T(n) \\
&\qquad + \mu^2 X(n)X^T(n)\tilde{W}(n)\tilde{W}^T(n)X(n)X^T(n)\Big] \\
&= E\Big[\tilde{W}(n)\tilde{W}^T(n) - 2\mu X(n)X^T(n)\tilde{W}(n)\tilde{W}^T(n) \\
&\qquad + \mu^2 X(n)X^T(n)\tilde{W}(n)\tilde{W}^T(n)X(n)X^T(n)\Big] \\
&= E\Big[\tilde{W}(n)\tilde{W}^T(n) - 2\mu X(n)X^T(n)\tilde{W}(n)\tilde{W}^T(n)\Big], (\mu >> 1) \\
&= E\Big[\Big(I - 2\mu X(n)X^T(n)\Big)\tilde{W}(n)\tilde{W}^T(n)\Big] \\
&= \Big(I - 2\mu E[X(n)X^T(n)]\Big)E[\tilde{W}(n)\tilde{W}^T(n)] \\
&= \Big(I - 2\mu R\Big)E[\tilde{W}(n)\tilde{W}^T(n)]
\end{aligned}
$$

2^{nd} term -
$$E\Big[\mu X(n)e_*(n)\tilde{W}(n)\tilde{W}^T(n)(I - \mu X(n)X^T(n))^T\Big]$$

$$
\begin{aligned}
&= E\Big[\mu X(n)e_*(n)\tilde{W}(n)\tilde{W}^T(n) \\
&\qquad - \mu^2 E\Big[\mu X(n)e_*(n)\tilde{W}(n)\tilde{W}^T(n)X(n)X^T(n)\Big] \\
&= \mu E\Big[X(n)e_*(n)\Big]E\Big[\tilde{W}(n)\tilde{W}^T(n)\Big] \\
&= \mu E\Big[X(n)[d(n) - X^T(n)W_*]\Big]E\Big[\tilde{W}(n)\tilde{W}^T(n)\Big] \\
&= \mu(r_{xd} - W_* R)E\Big[\tilde{W}(n)\tilde{W}^T(n)\Big] \\
&= 0
\end{aligned}
$$

Similarly 3^{rd} term
$$E\Big[(I - \mu X(n)X^T(n))\tilde{W}(n)\mu e_*^T(n)X^T(n)\Big] = 0$$

4^{th} term
$$E\Big[\mu^2 X(n)e_*(n)e_*^T(n)X^T(n)\Big]$$

$$
\begin{aligned}
&= E\Big[\mu^2 X(n)e_*^2(n)X^T(n)\Big] \\
&= \mu^2 E\Big[e_*^2(n)\Big]E\Big[X(n)X^T(n)\Big] \\
&= \mu^2 J_{min} R
\end{aligned}
$$

Let $R = U\Lambda U^T$ for some U, Λ,
such that Λ is a diagonal matrix and $U^T U = I$.

$P_u(n+1)$

$$= E\Big[U^T\tilde{W}(n+1)\tilde{W}^n(n+1)U\Big] + U^T\mu^2 J_{min}RU$$
$$= U^T\Big(I - 2\mu R\Big)E[\tilde{W}(n)\tilde{W}^T(n)]U + \mu^2 J_{min}U^T RU$$
$$= U^T\Big(UU^T - 2\mu U\Lambda U^T\Big)E[\tilde{W}(n)\tilde{W}^T(n)]U$$
$$+ \mu^2 J_{min}U^T U\Lambda U^T U$$
$$= U^T U\Big(I - 2\mu\Lambda\Big)U^T E[\tilde{W}(n)\tilde{W}^T(n)]U + \mu^2 J_{min}\Lambda$$
$$= \Big(I - 2\mu\Lambda\Big)P_u(n) + \mu^2 J_{min}\Lambda$$

Problem 2.10. Find the value of the cost function at infinity i.e. $J(\infty)$

Solution:

$$p_u^{ii}(\infty) = [1 - 2\mu\lambda_i]p_u^{ii}(\infty) + \mu^2 J_{min}\lambda_i$$
$$= \frac{\mu J_{min}}{2}$$
$$J_{ex}(\infty) = \sum_{i=1}^{M}\lambda_i p_u^{ii}(\infty)$$
$$= \sum_{i=1}^{M}\lambda_i\frac{\mu J_{min}}{2}$$
$$= \frac{\mu J_{min}}{2}\sum_{i=1}^{M}\lambda_i$$
$$= \frac{\mu J_{min}}{2}tr(R)$$
$$= \frac{\mu J_{min}}{2}M(SignalPower)$$
$$J(\infty) = J_{min} + \frac{\mu J_{min}}{2}\sum_{i=1}^{M}\lambda_i$$

Problem 2.11. How can you choose the value of μ from the convergence of both in mean and mean-square sense.

Solution:

1) For the convergence in mean, we require $0 < \mu < \dfrac{2}{M(SignalPower)}$
2) For the convergence in mean-square sense, we require $0 < \mu < \dfrac{1}{M(SignalPower)}$

$\therefore$ choose μ such that $0 < \mu < \dfrac{1}{M(SignalPower)}$

3 Convergence of the RLS Algorithm

3.1 Convergence in the Mean

Problem 3.1. Show that RLS converges in mean i.e. $E[W(n)] \to W_*$

Solution: From RLS algorithm

$$\hat{W}(n) = \phi^{-1}(n)\sum_{i=1}^{n}\lambda^{n-i}X(i)d^T(i) \qquad (3.1)$$

Where

$P(n) = \phi^{-1}(n) = [\sum_{i=1}^{n}\lambda^{n-i}X(i)X^T(i)]^{-1}$

$\therefore\ \hat{W}(n) = [\sum_{i=1}^{n}\lambda^{n-i}X(i)X^T(i)]^{-1}[\sum_{i=1}^{n}\lambda^{n-i}X(i)d(i)]$

δ I is added to make sure that inverse exists. Weiner optimal error is

$$e_*(n) = d(n) - W_*X(n)$$
$$d(n) = e_*(n) - W_*X(n)$$

W(n)

$$= [\delta I + \sum_{i=1}^{n}\lambda^{n-i}X(i)X^T(i)]^{-1}[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(n) - W_*(n)X(n))]$$
$$= [\delta I + \sum_{i=1}^{n}\lambda^{n-i}X(i)X^T(i)]^{-1}[\delta IW_* + \sum_{i=1}^{n}\lambda^{n-i}X(i)X^T(i)^{-1}W_* + \sum$$
$$= W_* + [\delta I + \sum_{i=1}^{n}\lambda^{n-i}X(i)X^T(i)]^{-1}[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(n) - \delta IW_*]$$

We know that
$R = \frac{1}{n}\sum_{i=1}^{n}\lambda^{n-i}X(i)X^T(i)$
$\phi(n) = \delta I + \sum_{i=1}^{n}\lambda^{n-i}X(i)X^T(i)$
For $\lambda = 1$

$$R = \frac{\phi(n)}{n} - \frac{\delta I}{n}$$
$$[\phi[n]]^{-1} = \frac{1}{n}R^{-1} \qquad (3.2)$$

$$W(n) = W_* + \frac{1}{n}R^{-1}\frac{1}{n}R^{-1}\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(n) - \delta IW_*\Big] \qquad (3.3)$$

$$E[W(n)] = E\Big[W_* + \frac{1}{n}R^{-1}\frac{1}{n}R^{-1}\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(n) - \delta I W_*\Big]\Big]$$
$$= W_* + \frac{1}{n}R^{-1}E\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(n)\Big] - \frac{1}{n}R^{-1}E[\delta I W_*]$$
$$= W_* + \frac{1}{n}R^{-1}\sum_{i=1}^{n}\lambda^{n-i}E[X(i)(e_*(n)] - \frac{1}{n}R^{-1}\delta I W_*$$
$$= W_* - \frac{\delta I}{n}R^{-1}W_*$$

$$\lim_{n\to\infty} E[W(n)] = W_* \tag{3.4}$$

3.2 Convergence in Mean-Square sense

Problem 3.2. Using Equation 3.3 Show that RLS converges in mean square sense i.e.

$$\lim_{n\to\infty} E[(W(n) - W_*)(W(n) - W_*)^T] = 0 \tag{3.5}$$

Solution: Let $S(n) = E[(W(n) - W_*)(W(n) - W_*)^T]$
From Equation 3.3
$W(n) - W_* = \frac{1}{n}R^{-1}\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(n) - \delta I W_*\Big]$

S(n)

$$= E\Big\{\Big(\frac{1}{n}R^{-1}\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(n) - \delta I W_*\Big]\Big)\Big(\frac{1}{n}R^{-1}\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(n) - \delta I W_*\Big]\Big)^T\Big\}$$
$$= E\Big\{\frac{\delta^2 I R^{-1}W_*W_*^T(R^{-1})^T}{n^2} - \frac{\delta I}{n^2}R^{-1}W_*\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(i)\Big](R^{-1})^T - \Big(\frac{\delta I}{n^2}R^{-1}W_*\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(i)\Big](R^{-1})^T\Big)^T + \frac{1}{n^2}R^{-1}\Big(\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(i)\Big)\Big(\sum_{i=1}^{n}\lambda^{n-i}X(i)e_*(i)\Big)^T(R^{-1})^T\Big)\Big\}$$
$$= E\Big\{\frac{\delta^2 I R^{-1}W_*W_*^T(R^{-1})^T}{n^2} - \frac{\delta I}{n^2}R^{-1}W_*\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(i)\Big](R^{-1})^T - \Big(\frac{\delta I}{n^2}R^{-1}W_*\Big[\sum_{i=1}^{n}\lambda^{n-i}X(i)(e_*(i)\Big](R^{-1})^T\Big)^T + \frac{1}{n^2}R^{-1}\Big(\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda^{n-i}\lambda^{n-i}\lambda^{n-j}X(i)e_*(i)e_*^T(j)X^T(j)\Big)(R^{-1})^T\Big\}$$

Assumptions
$X(i) \perp\!\!\!\perp X(j), d(i) \perp\!\!\!\perp d(jd(i), \perp\!\!\!\perp X(j)$
$e_*(i) = d(i) - X^T(i)W_*$
The 1st term is deterministic and from the above assumptions the Expectations 2nd and 3rd become zero.

$$E\Big[X(i)e_*(i)e_*^T(j)X^T(j)\Big] = E[X(i)X^T(j)]E[e_*(i)e_*^T(j)] = RJ_{min}$$

$$S(n) = \frac{\delta^2 I R^{-1}W_*W_*^T(R^{-1})^T}{n^2} + \frac{1}{n^2}R^{-1}RJ_{min}(R^{-1})^T\sum_{i=1}^{n}\Big(\lambda^{n-i}\Big)^2$$

$$\lim_{n\to\infty} S(n) = \lim_{n\to\infty} E\Big[(W(n) - W_*)(W(n) - W_*)^T\Big] = 0 \tag{3.6}$$