Second-order Convergence Analysis of the LMS Algorithm and Misadjustment Error*

Douglas L. Jones

Convergence of the mean (first-order analysis) is insufficient to guarantee desirable behavior of the LMS algorithm; the variance could still be infinite. It is important to show that the variance of the filter coefficients is finite, and to determine how close the average squared error is to the minimum possible error using an exact Wiener filter.

$$\begin{aligned} E\left[{\epsilon_k}^2\right] &= E\left[\left(d_k - {W^k}^T X^k\right)^2\right] \\ &= E\left[{d_k}^2 - 2d_k {X^k}^T W^k - {W^k}^T X^k {X^k}^T W^k\right] \\ &= r_{\mathrm{dd}}(0) - 2{W^k}^T P + {W^k}^T R W^k \end{aligned} \quad (1)$$

The minimum error is obtained using the Wiener filter

$$W_{\mathrm{opt}} = R^{-1} P$$

$$\begin{aligned} {\epsilon_{\min}}^2 &= E\left[\epsilon^2\right] \\ &= \left(r_{\mathrm{dd}}(0) - 2P^T R^{-1} P + P^T R^{-1} R R^{-1} P\right) \\ &= r_{\mathrm{dd}}(0) - P^T R^{-1} P \end{aligned} \quad (2)$$

To analyze the average error in LMS, write (1) in terms of $V' = Q\left[W - W_{\mathrm{opt}}\right]$, where $Q^{-1} \Lambda Q = R$

$$E\left[{\epsilon_k}^2\right] = r_{\mathrm{dd}}(0) - 2{W^k}^T P + {W^k}^T R W^k + \left(-\left({W^k}^T R W_{\mathrm{opt}}\right)\right) - {W_{\mathrm{opt}}}^T R W^k + {W_{\mathrm{opt}}}^T R W_{\mathrm{opt}} + {W^k}^T R W_{\mathrm{opt}} + {W_{\mathrm{opt}}}^T R W^k - {W_{\mathrm{opt}}}^T R W_{\mathrm{opt}} = r_{\mathrm{dd}}(0) + {V^k}^T R V^k - P^T R^{-1} P = {\epsilon_{\min}}^2 + {V^k}^T R V^k = {\epsilon_{\min}}^2 + {V^k}^T Q^{-1} Q R Q^{-1} Q V^k = {\epsilon_{\min}}^2 + {V'^k}^T \Lambda V'^k \quad (3)$$

$$E\left[{\epsilon_k}^2\right] = {\epsilon_{\min}}^2 + \sum_{j=0}^{N-1} \lambda_j E\left[{v_j'^k}^2\right]$$

So we need to know $E\left[{v_j'^k}^2\right]$, which are the diagonal elements of the covariance matrix of V'^k, or $E\left[V'^k {V'^k}^T\right]$.

*Version 1.2: May 11, 2004 11:04 am -0500
†http://creativecommons.org/licenses/by/1.0

From the LMS update equation

$$W^{k+1} = W^k + 2\mu\epsilon_k X^k$$

we get

$$V^{\text{'k}+1} = W^{\text{'k}} + 2\mu\epsilon_k Q X^k$$

$$\begin{array}{rcl} \mathscr{V}^{k+1} & = & E\left[V^{\text{'}k+1}{V^{\text{'}k+1}}^T\right] \\ & = & E\left[4\mu^2{\epsilon_k}^2 Q X^k {X^k}^T Q^T\right] \\ & = & \mathscr{V}^k + 2\mu\left(\epsilon_k Q X^k {V^{\text{'k}}}^T\right) + 2\mu\left(\epsilon_k V^{\text{'k}} {X^k}^T Q^T\right) + 4\mu^2 E\left[{\epsilon_k}^2 Q X^k {X^k}^T Q^T\right] \end{array} \quad (4)$$

Note that

$$\epsilon_k = d_k - {W^k}^T X^k = d_k - {W_{\text{opt}}}^T - {V^{\text{'k}}}^T Q X^k$$

so

$$\begin{array}{rcl} E\left[\epsilon_k Q X^k {V^{\text{'k}}}^T\right] & = & E\left[d_k Q X^k {V^{\text{'k}}}^T - {W_{\text{opt}}}^T X^k Q X^k {V^{\text{'k}}}^T - {V^{\text{'k}}}^T Q X^k {V^{\text{'k}}}^T\right] \\ & = & 0 + 0 - \left(Q X^k {X^k}^T Q^T V^{\text{'k}} {V^{\text{'k}}}^T\right) \\ & = & -\left(Q E\left[X^k {X^k}^T\right] Q^T E\left[V^{\text{'k}} {V^{\text{'k}}}^T\right]\right) \\ & = & -\left(\Lambda \mathscr{V}^k\right) \end{array} \quad (5)$$

Note that the Patently False independence Assumption was invoked here.

To analyze $E\left[{\epsilon_k}^2 Q X^k {X^k}^T Q^T\right]$, we make yet another obviously false assumptioon that ${\epsilon_k}^2$ and X^k are statistically independent. This is obviously false, since $\epsilon_k = d_k - {W^k}^T X^k$. Otherwise, we get 4th-order terms in X in the product. These can be dealt with, at the expense of a more complicated analysis, if a particular type of distribution (such as Gaussian) is assumed. See, for example *Gardner*[1]. A questionable justification for this assumption is that as $W^k \simeq W_{\text{opt}}$, W^k becomes uncorrelated with X^k (if we invoke the original independence assumption), which tends to randomize the error signal relative to X^k. With this assumption,

$$E\left[{\epsilon_k}^2 Q X^k {X^k}^T Q^T\right] = E\left[{\epsilon_k}^2\right] E\left[Q X^k {X^k}^T Q^T\right] = E\left[{\epsilon_k}^2\right] \Lambda$$

Now

$${\epsilon_k}^2 = {\epsilon_{\min}}^2 + {V^{\text{'k}}}^T \Lambda V^{\text{'k}}$$

so

$$\begin{array}{rcl} E\left[{\epsilon_k}^2\right] & = & {\epsilon_{\min}}^2 + E\left[\sum_j \lambda_j {V_j^{\text{'k}}}^2\right] \\ & = & {\epsilon_{\min}}^2 + \sum_j \lambda_j \mathscr{V}_{\text{jj}}^k \end{array} \quad (6)$$

Thus, (4) becomes

$$\mathscr{V}^{k+1} = I\mathscr{V}^k + 4\mu^2 \sum_j \lambda_j \mathscr{V}_{\text{jj}}^k \Lambda + 4\mu^2 {\epsilon_{\min}}^2 \Lambda \quad (7)$$

Now **if** this system is stable and converges, it converges to $\mathscr{V}^\infty = \mathscr{V}^{\infty+1}$

$$\Rightarrow \left(4\mu\Lambda\mathscr{V}^\infty = 4\mu^2\left(\sum_j \lambda_j \mathscr{V}_{\text{jj}} + {\epsilon_{\min}}^2\right)\Lambda\right)$$

$$\Rightarrow \left(\mathscr{V}^\infty = \mu\left(\sum_j \lambda_j \mathscr{V}_{\text{jj}} + {\epsilon_{\min}}^2\right) I\right)$$

So it is a diagonal matrix with all elements on the diagonal equal:

Then

$$\mathcal{V}_{\mathrm{ii}}^{\infty} = \mu \left(\mathcal{V}_{\mathrm{ii}}^{\infty} \sum_j \lambda_j + {\epsilon_{\min}}^2 \right)$$

$$\mathcal{V}_{\mathrm{ii}}^{\infty} \left(1 - \mu \sum_j \lambda_j \right) = \mu {\epsilon_{\min}}^2$$

$$\mathcal{V}_{\mathrm{ii}}^{\infty} = \frac{\mu {\epsilon_{\min}}^2}{1 - \mu \sum_j \lambda_j}$$

Thus the error in the LMS adaptive filter after convergence is

$$\begin{aligned} E\left[{\epsilon_\infty}^2\right] &= {\epsilon_{\min}}^2 + E\left[V'^{\infty} \lambda V'^{\infty}\right] \\ &= {\epsilon_{\min}}^2 + \frac{\mu {\epsilon_{\min}}^2 \sum_j \lambda_j}{1 - \mu \sum_j \lambda_j} \\ &= {\epsilon_{\min}}^2 \frac{1}{1 - \mu \sum_j \lambda_j} \\ &= {\epsilon_{\min}}^2 \frac{1}{1 - \mu \mathrm{tr}(R)} \\ &= {\epsilon_{\min}}^2 \frac{1}{1 - \mu r_{\mathrm{xx}}(0) N} \end{aligned} \quad (8)$$

$$E\left[{\epsilon_\infty}^2\right] = {\epsilon_{\min}}^2 \frac{1}{1 - \mu N {\sigma_x}^2} \quad (9)$$

$1 - \mu N {\sigma_x}^2$ is called the **misadjustment factor**. Oftern, one chooses μ to select a desired misadjustment factor, such as an error 10% higher than the Wiener filter error.

1 2nd-Order Convergence (Stability)

To determine the range for μ for which (7) converges, we must determine the μ for which the matrix difference equation converges.

$$\mathcal{V}^{k+1} = I \mathcal{V}^k + 4\mu^2 \sum_j \lambda_j \mathcal{V}_{\mathrm{jj}}^k \Lambda + 4\mu^2 {\epsilon_{\min}}^2 \Lambda$$

The off-diagonal elements each evolve independently according to $\mathcal{V}_{\mathrm{ij}}^{k+1} = 1 - 4\mu\lambda_i \mathcal{V}_{\mathrm{ij}}^k$ These terms will decay to zero if $\forall i : (4\mu\lambda_i < 2)$, or $\mu < \frac{1}{2\lambda_{\max}}$

The diagonal terms evolve according to

$$\mathcal{V}_{\mathrm{ii}}^{k+1} = 1 \mathcal{V}_{\mathrm{ii}}^k + 4\mu^2 \lambda_i \sum_j \lambda_j \mathcal{V}_{\mathrm{jj}}^k + 4\mu^2 {\epsilon_{\min}}^2 \lambda_i$$

For the homoegeneous equation

$$\mathcal{V}_{\mathrm{ii}}^{k+1} = 1 \mathcal{V}_{\mathrm{ii}}^k + 4\mu^2 \lambda_i \sum_j \lambda_j \mathcal{V}_{\mathrm{jj}}^k$$

for $1 - 4\mu\lambda_i$ positive,

$$\mathcal{V}_{\mathrm{ii}}^{k+1} \leq 1 \mathcal{V}_{\mathrm{iimax}}^k + 4\mu^2 \lambda_i \sum_j \lambda_j \mathcal{V}_{\mathrm{jjmax}}^k = \left(1 - 4\mu\lambda_i + 4\mu^2 \lambda_i \sum_j \lambda_j \right) \mathcal{V}_{\mathrm{jjmax}}^k \quad (10)$$

$\mathcal{V}_{\mathrm{ii}}^{k+1}$ will be strictly less than $\mathcal{V}_{\mathrm{jjmax}}^k$ for

$$1 - 4\mu\lambda_i + 4\mu^2 \lambda_i \sum_j \lambda_j < 1$$

or

$$4\mu^2\lambda_i \sum_j \lambda_j < 4\mu\lambda_i$$

or

$$\begin{aligned} \mu < \frac{1}{\sum_j \lambda_j} &= \frac{1}{\mathrm{tr}(R)} \\ &= \frac{1}{N r_{\mathrm{xx}}(0)} \\ &= \frac{1}{N {\sigma_x}^2} \end{aligned} \quad (11)$$

This is a more rigorous bound than the first-order bounds. Ofter engineers choose μ a few times smaller than this, since more rigorous analyses yield a slightly smaller bound. $\mu = \frac{\mu}{3N{\sigma_x}^2}$ is derived in some analyses assuming Gaussian x_k, d_k.

References

[1] W.A. Gardner. Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique. *Signal Processing*, 6:113–133, 1984.