Stability and Performance Analysis of the LMS Algorithm

Convergence in the Mean Square

The stronger condition for "convergence" of an LMS filter is to require convergence in the mean square, i.e.

$$\lim_{n\to\infty} J(n) = c \geq J_{\min}$$

where c is a small constant.

This work was first published in the classic paper by A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, and Signal Proc., Feb. 1985*. These lecture notes use the elegant derivation presented by M. Rupp, "The behavior of LMS and NLMS in the presence of spherically invariant processes," *IEEE Trans. on Signal Proc., Mar. 1993*. We note that the mean square convergence analysis in Haykin, *Adaptive Filter Theory 3rd E.* is not correct and specifically that the bounds in (9.83) and (9.92) are incorrect.

Outline of Analysis

1. We show $J(n) = J_{\min} + \mathrm{tr}[\mathbf{RK}(n)]$ where the "excess MSE" is defined as $J_{ex} = \mathrm{tr}[\mathbf{RK}(n)]$ with $\mathbf{K}(n) = E[\varepsilon(n)\varepsilon^H(n)]$ defined as the weight error vector (or misalignment) covariance matrix. The goal is to determine conditions which force J_{ex} to be a small constant (we cannot do anything about $J_{\min}$).

2. We examine the transient behavior of $\mathbf{K}(n)$ by expressing $\mathbf{K}(n+1)$ in terms of $\mathbf{K}(n)$ and $\mathbf{R}$ to aid us in step 3.

3. We analyze requirements on μ to guarantee that J_{ex} goes to a small constant and hence $J(n)$ goes to a constant, i.e. convergence in the mean square.

---

1. We can express the estimation error from the filter, $\hat{\mathbf{w}}(n)$ as

$$\begin{aligned} e(n) &= \delta(\nu) - \hat{\boldsymbol{\omega}}^H(\nu)\boldsymbol{\upsilon}(\nu) \\ &= \delta(\nu) - \boldsymbol{\omega}_{\mathrm{o}}^H(\nu)\boldsymbol{\upsilon}(\nu) - \varepsilon^H(\nu)\boldsymbol{\upsilon}(\nu) \\ &= \varepsilon_o(\nu) - \varepsilon^H(\nu)\boldsymbol{\upsilon}(\nu) \end{aligned}$$

and the MSE as

$$\begin{aligned} J(n) &= E[e(n)e^*(n)] \\ &= E\{[e_{\mathrm{o}}(n) - \varepsilon^H(n)\mathbf{u}(n)][e_{\mathrm{o}}(n) - \varepsilon^H(n)\mathbf{u}(n)]^*\} \\ &= E[e_{\mathrm{o}}(n)e_{\mathrm{o}}^*(n)] - E[e_{\mathrm{o}}^*(n)\varepsilon^H(n)\mathbf{u}(n)] - E[e_{\mathrm{o}}(n)\mathbf{u}^H(n)\varepsilon(n)] + E[\varepsilon^H(n)\mathbf{u}(n)\mathbf{u}^H(n)\varepsilon(n)] \\ &= E[e_{\mathrm{o}}(n)e_{\mathrm{o}}^*(n)] - E[\varepsilon^H(n)]E[e_{\mathrm{o}}^*(n)\mathbf{u}(n)] - E[e_{\mathrm{o}}(n)\mathbf{u}^H(n)]E[\varepsilon(n)] + E[\varepsilon^H(n)\mathbf{u}(n)\mathbf{u}^H(n)\varepsilon(n)] \\ &= J_{\min} + E[\varepsilon^H(n)\mathbf{u}(n)\mathbf{u}^H(n)\varepsilon(n)] \end{aligned}$$

where in the 4th line we have invoked the independence assumption [ε(n) is independent of $\mathbf{u}(n)$ and $d(n)$ and hence $e(n)$] and in the 5th line we have invoked the principle of orthogonality. Since the last term on the right hand side (RHS) on the final line is a scalar [and scalar = tr(scalar)] we have

$$\begin{aligned} E[\varepsilon^H(n)\mathbf{u}(n)\mathbf{u}^H(n)\varepsilon(n)] &= E\{\mathrm{tr}[\varepsilon^H(n)\mathbf{u}(n)\mathbf{u}^H(n)\varepsilon(n)]\} \\ &= \mathrm{tr}\{E[\mathbf{u}(n)\mathbf{u}^H(n)\varepsilon(n)\varepsilon^H(n)]\} \\ &= \mathrm{tr}\{E[\mathbf{u}(n)\mathbf{u}^H(n)]E[\varepsilon(n)\varepsilon^H(n)]\} \\ &= \mathrm{tr}[\mathbf{RK}(n)] \end{aligned}$$

where again we have invoked the independence assumption in the third line. Thus our expression for the MSE is

$$\boxed{J(n) = J_{\min} + \mathrm{tr}[\mathbf{RK}(n)]}$$

We note this expression consists of two parts: the MMSE $J_{\min}$ and a component which depends on the transient behavior of the weight-error covariance matrix $\mathbf{K}(n)$.

---

2. In order to analyze convergence of the MSE, we examine the transient behavior of the weight-error covariance matrix, $\mathbf{K}(n)$. This analysis will provide insight into the transient behavior of the MSE of the LMS algorithm. We have

$$\varepsilon(n+1) = [\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}^H(n)]\varepsilon(n) + \mu\mathbf{u}(n)e_o^*(n)$$

which yields (see handout) the matrix difference equation

$$\boxed{\begin{aligned} \mathbf{K}(n+1) &= E[\varepsilon(n+1)\varepsilon^H(n+1)] \\ &= \mathbf{K}(n) - \mu\mathbf{RK}(n) - \mu\mathbf{K}(n)\mathbf{R} + \mu^2[\alpha\mathbf{RK}(n)\mathbf{R} + \mathbf{R}\,\mathrm{tr}(\mathbf{RK}(n))] + \mu^2 J_{\min}\mathbf{R} \end{aligned}} \quad \text{[Rupp (3.13)]}$$

where $\alpha = 2$ if $\mathbf{u}(n)$ is real-valued and $\alpha = 1$ if $\mathbf{u}(n)$ is complex-valued—a result of the Gaussian moment factoring theorem) [compare to Haykin, (9.62)].

Clearly $\mathbf{K}(n)$ does not go to $\mathbf{0}$ as n goes to ∞ (if we backsolve) due to the excitation caused by $\mu^2 J_{\min}\mathbf{R}$.

Note that it can be shown by induction and some positive definite reasoning, that

$$\mathbf{K}(n+1) > \mathbf{0}$$

positive definite).

---

3. We want to determine conditions that will force J_{ex} to a small constant. Since $\mathbf{RK}(n) > \mathbf{0}$ (product of positive definite is positive definite), $J(n) > J_{\min}$ and so the MSE $J(n)$, is always in excess of the MMSE $J_{\min}$. This difference was defined earlier as the excess MSE and is given by

$$\begin{aligned} J_{ex}(n) &= J(n) - J_{\min} \\ &= \mathrm{tr}[\mathbf{RK}(n)] \\ &= \mathrm{tr}[\mathbf{Q}^H\mathbf{RQQ}^H\mathbf{K}(n)\mathbf{Q}] \\ &= \mathrm{tr}[\Lambda\mathbf{X}(n)] \end{aligned}$$

where

$$\mathbf{X}(n) = \mathbf{\Theta}^H\mathbf{K}(\nu)\mathbf{\Theta}$$

and is not diagonal in general. Finally we note

$$\boxed{\begin{aligned} J_{ex}(n) &= \sum_{\iota=1}^{M} \lambda_\iota \xi_\iota(\nu) \\ &= \lambda^T \boldsymbol{\xi}(\nu) \end{aligned}}$$

where

$$\begin{aligned} \mathrm{diag}[\mathbf{X}(n)] &= \boldsymbol{\xi}(\nu) \\ &= [\xi_1(\nu) \;\; \Lambda \;\; \xi_M(\nu)]^T \end{aligned}$$

and

$$\begin{aligned} \mathrm{diag}(\Lambda) &= \lambda \\ &= [\lambda_1 \;\; \Lambda \;\; \lambda_M]^T. \end{aligned}$$

Let's get a handle on $\mathbf{X}(n)$ so as to understand $\mathbf{x}(n)$ and hence J_{ex}. Using the result in 2) we have

$$\begin{aligned} \mathbf{X}(n+1) &= \mathbf{Q}^H\mathbf{K}(n+1)\mathbf{Q} \\ &= \mathbf{Q}^H\mathbf{K}(n)\mathbf{Q} - \mu\mathbf{Q}^H\mathbf{RQQ}^H\mathbf{K}(n)\mathbf{Q} - \mu\mathbf{Q}^H\mathbf{K}(n)\mathbf{QQ}^H\mathbf{RQ} + . \\ &\quad \mu^2\mathbf{Q}^H\{\alpha\mathbf{RK}(n)\mathbf{R} + \mathbf{R}\,\mathrm{tr}[\mathbf{RK}(n)]\}\mathbf{Q} + \mu^2 J_{\min}\mathbf{Q}^H\mathbf{RQ} \end{aligned}$$

We note that for any $\mathbf{A}$

$$\mathbf{Q}^H\ \mathrm{tr}[\mathbf{A}]\mathbf{Q} = \mathrm{tr}[\mathbf{Q}^H\mathbf{AQ}].$$

Hence,

$$\begin{aligned} \mu^2\mathbf{Q}^H\{\alpha\mathbf{RK}(n)\mathbf{R} + \mathbf{R}\,\mathrm{tr}[\mathbf{RK}(n)]\}\mathbf{Q} &= \mu^2\mathbf{Q}^H\{\alpha\mathbf{RQQ}^H\mathbf{K}(n)\mathbf{QQ}^H\mathbf{R} + \mathbf{RQQ}^H\ \mathrm{tr}[\mathbf{RK}(n)]\}\mathbf{Q} \\ &= \alpha\mu^2\mathbf{Q}^H\mathbf{RQQ}^H\mathbf{K}(n)\mathbf{QQ}^H\mathbf{RQ} + \mu^2\mathbf{Q}^H\mathbf{RQ}\,\mathrm{tr}[\mathbf{Q}^H\mathbf{RQQ}^H\mathbf{K}(n)\mathbf{Q}]. \\ &= \alpha\mu^2\Lambda\mathbf{X}(n)\Lambda + \mu^2\Lambda\,\mathrm{tr}[\Lambda\mathbf{X}(n)] \end{aligned}$$

Substituting we have

$$\mathbf{X}(n+1) = \Xi(\nu) - \mu\Lambda\Xi(\nu) - \mu\Xi(\nu)\Lambda + \alpha\mu^2\Lambda\Xi(\nu)\Lambda + \mu^2\Lambda\tau\rho[\Lambda\Xi(\nu)] + \mu^2\vartheta_{\mu\iota\nu}\Lambda \qquad \text{[Rupp, (3.15)]}$$

[compare to Haykin, (9.73)]

Since $J_{\mathrm{ex}}(n)$ depends on $\mathbf{x}(n)$, we need only look at the diagonal terms of $\mathbf{X}(n+1)$, expressed recursively as

$$\mathbf{x}(n+1) = \mathbf{B}\boldsymbol{\xi}(\nu) + \mu^2\vartheta_{\mu\iota\nu}\lambda \qquad \text{[Rupp, (3.16)]}$$

where

$$b_{ij} = \begin{cases} 1 - 2\mu\lambda_\iota + (\alpha+1)\mu^2\lambda_\iota^2, & \iota = \varphi \\ \mu^2\lambda_\iota\lambda_\varphi & \iota \neq \varphi \end{cases}. \qquad \text{[Rupp, (3.17)]}$$

[compare to Haykin, (9.76)]

In order for $\mathbf{x}(n)$ (and hence J_{ex}) to converge, eigenvalues of $\mathbf{B}$ must lie within the unit circle, i.e. the spectral radius, $\rho(\mathbf{B}) < 1$. This eigenvalue condition will lead to bounds on μ according to b_{ij}.