Least-Mean-Square Algorithm

Stability and Performance Analysis of the LMS Algorithm

Independence Assumption

In order to greatly simplify our analysis, we introduce a fundamental assumption (known as the independence assumption) consisting of four parts:

1) Input vectors $\mathbf{u}(1), \mathbf{u}(2), \ldots, \mathbf{u}(n)$ constitute a sequence of statistically independent vectors (in a temporal setting, these are actually statistically dependent)
2) Input vector $\mathbf{u}(n)$ is statistically independent of all previous samples of the desired response, $d(1), d(2), \ldots, d(n-1)$.
3) Desired response, $d(n)$ is dependent on $\mathbf{u}(n)$ but statistically independent of $d(1), d(2), \ldots, d(n-1)$.
4) Input vector $\mathbf{u}(n)$ and the desired response $d(n)$ consist of mutually Gaussian-distributed random variables for all n.

We note that based on the LMS update equation and assumptions 1) and 2), $\hat{\mathbf{w}}(n+1)$, and therefore the misalignment vector, $\varepsilon(n+1) = \mathbf{w}(n+1) - \mathbf{w}_o$, are independent of both $\mathbf{u}(n+1)$ and $d(n+1)$, i.e.

$$E[\mathbf{u}(n)\varepsilon(n)] = E[\mathbf{u}(n)]E[\varepsilon(n)].$$

We have emphasized statistical independence of samples, rather than unorrelatedness. It is indeeed true that independent samlpes are equivalent to uncorrelated samples whenever the samples are Gaussian distributed. Thus assumptions 1) and 2) are equivalent to the following conditions of uncorrelatedness:

$$E[\mathbf{u}(n)\mathbf{u}^H(k)] = \mathbf{0}, \quad k = 0, 1, \ldots, n-1$$
$$E[\mathbf{u}(n)d^*(k)] = \mathbf{0}, \quad k = 0, 1, \ldots, n-1$$

Convergence of the Mean

To prove convergence of the mean, we wish to show that on average, the weight vector goes to the Wiener filter

$$\lim_{n\to\infty} E[\hat{\mathbf{w}}(n)] = \mathbf{w}_o$$

or equivalently, for the misalignment vector, $\varepsilon(n+1) = \mathbf{w}(n+1) - \mathbf{w}_o$

$$\lim_{n\to\infty} E[\varepsilon(n)] = \mathbf{0}.$$

Beginning with the LMS update algorithm and subtracting the Wiener (optimal) filter from both sides we have

$$\hat{\mathbf{w}}(n+1) - \boldsymbol{\omega}_o = \hat{\boldsymbol{\omega}}(\nu) + \mu\boldsymbol{\upsilon}(\nu)[\delta^*(\nu) - \boldsymbol{\upsilon}^H(\nu)\hat{\boldsymbol{\omega}}(\nu)] - \boldsymbol{\omega}_o$$

or

$$\begin{aligned}
\varepsilon(n+1) &= \varepsilon(n) + \mu\mathbf{u}(n)d^*(n) - \mu\mathbf{u}(n)\mathbf{u}^H(n)\hat{\mathbf{w}}(n) \\
&= \varepsilon(n) + \mu\mathbf{u}(n)d^*(n) - \mu\mathbf{u}(n)\mathbf{u}^H(n)[\varepsilon(n) + \mathbf{w}_o] \\
&= [\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}^H(n)]\varepsilon(n) + \mu\mathbf{u}(n)[d^*(n) - \mathbf{u}^H(n)\mathbf{w}_o] \\
&= [\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}^H(n)]\varepsilon(n) + \mu\mathbf{u}(n)e_o^*(n)
\end{aligned}$$

where

$$e_o(n) = d(n) - \mathbf{w}_o^H \mathbf{u}(n)$$

is the estimation error produced by the Wiener filter. As a side note, we assume the first-order moment is zero

$$E[e_o(n)] = 0$$

and the second moment is the MSE associated with the Wiener solution

$$E[e_o(n)\varepsilon_o^*(\nu)] = \vartheta_{\mu\iota\nu} \ .$$

The expected misalignment (coefficient error) is

$$E[\varepsilon(\nu+1)] = E\{[\mathbf{I} - \mu\boldsymbol{\upsilon}(\nu)\boldsymbol{\upsilon}^H(\nu)]\varepsilon(\nu)\} + E[\mu\boldsymbol{\upsilon}(\nu)\varepsilon_o^*(\nu)] \ .$$

Invoking (*) of the independence assumption and assuming that $e_o(n)$ is orthogonal to $\mathbf{u}(n)$ (orthogonality principle) we have

$$\begin{aligned} E[\varepsilon(\nu+1)] &= \mathbf{I} - \mu E[\boldsymbol{\upsilon}(\nu)\boldsymbol{\upsilon}^H(\nu)]E[\varepsilon(\nu)] \\ &= (\mathbf{I} - \mu\mathbf{P})E[\varepsilon(\nu)] \\ &= (\mathbf{I} - \mu\mathbf{P})^{\nu+1} E[\varepsilon(0)] \end{aligned} \ .$$

We solved this exact equation in the SD analysis. We premultiply by the orthogonal matrix, **Q** which diagonalizes **R**

$\mathbf{Q}^H E[\varepsilon(\nu+1)] = \mathbf{\Theta}^H (\mathbf{I} - \mu\mathbf{P})^{\nu+1} E[\varepsilon(0)]$

which can be shown to yield (using the fact that **Q** is unitary and $\mathbf{Q}^H\mathbf{R}\mathbf{Q} = \Lambda$)

$$E[\omega(\nu+1)] = (\mathbf{I} - \mu\Lambda)^{\nu+1} E[\omega(0)]$$

where

$$\omega(n) = \mathbf{Q}^H \varepsilon(n) \; .$$

If we have (as in SD)

$$0 < \mu < \frac{2}{\lambda_{\mu\alpha\xi}}$$

then

$$E[\omega(\nu+1)] \to \mathbf{0}$$

or since the unitary transform does not change a vector's length,

$$E[\varepsilon(n+1)] \to \mathbf{0} \; .$$

Remarks: If $0 < \mu < \frac{2}{\lambda_{\mu\alpha\xi}}$, we expect convergence (convergence of the mean) with the LMS algorithm, i.e.

$$\lim_{n\to\infty} E[\varepsilon(\nu)] = 0$$

or equivalently

$$\lim_{n\to\infty} E[\hat{\boldsymbol{\omega}}(\nu)] = \boldsymbol{\omega}_o \; .$$

Since any sequence of zero-mean random variables converges in this sense this isn't too useful and does not guarantee stability of the complete algorithm.

Clearly, if $\lim_{n\to\infty} \hat{\boldsymbol{\omega}}(\nu) \neq \boldsymbol{\omega}_o$ (note no expectation here) we don't get $\lim_{n\to\infty} \vartheta(\nu) \approx \vartheta_{\mu\iota\nu}$ which is what we are ultimately after. Thus we'd now like to analyze conditions for "convergence in the mean square," i.e.

$$\begin{aligned} \lim_{n\to\infty} E[\varepsilon(\nu)\varepsilon^*(\nu)] &= \lambda\iota\mu_{\nu\to\infty}\, \vartheta(\nu) \\ &= \chi o\nu\sigma\tau\alpha\nu \end{aligned} \; .$$