Convergence Analysis of the RLS Algorithm

For this analysis we'll assume that the desired signal $d(n)$ can be modeled as

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

where $\mathbf{w}_o$ is the coefficient vector of the model (or solution to the LS problem) and $e_o(n)$ is the measurement error (white, zero-mean, σ^2 variance).

Convergence of the Mean

For convergence of the mean we need to show

$$E[\hat{\mathbf{w}}(n)] = \mathbf{w}_o \,.$$

We begin by assuming $\lambda = 1$ to get

$$\begin{aligned} \Phi(n) &= \sum_{i=1}^{n} \mathbf{u}(i)\mathbf{u}^H(i) \\ \mathbf{z}(n) &= \sum_{i=1}^{n} \mathbf{u}(i) d^*(i) \end{aligned} \,.$$

Next, assuming $n \geq M$, i.e. transient period is over, we have as the solution to the normal equations

$$\begin{aligned} \hat{\mathbf{w}}(n) &= \Phi^{-1}(n)\mathbf{z}(n) \\ &= \Phi^{-1}(n)\sum_{i=1}^{n} \mathbf{u}(i) d^*(i) \\ &= \Phi^{-1}(n)\sum_{i=1}^{n} \mathbf{u}(i)[\mathbf{u}^H(i)\mathbf{w}_o + e_o^*(i)] \\ &= \Phi^{-1}(n)\sum_{i=1}^{n} \mathbf{u}(i)\mathbf{u}^H(i)\mathbf{w}_o + \Phi^{-1}(n)\sum_{i=1}^{n} \mathbf{u}(i) e_o^*(i) \\ &= \mathbf{w}_o + \Phi^{-1}(n)\sum_{i=1}^{n} \mathbf{u}(i) e_o^*(i) \end{aligned} \,.$$

For the discrete random variable case, we define the conditional expectation as

$$E[X|Y] = \sum_{x} x P\{X = x | Y = y\}$$

for all values of y where $P_Y\{y\} > 0$. It can be proved that

$$E[X] = E[E[X|Y]] \,.$$

Employing the conditional expectation property, we have

$$
\begin{aligned}
E[\hat{\mathbf{w}}(n)] &= E\left[\boldsymbol{\omega}_o + \Phi^{-1}(\nu)\sum_{\iota=1}^{\nu}\boldsymbol{\upsilon}(\iota)\varepsilon_o^*(\iota)\right] \\
&= \boldsymbol{\omega}_o + E\left[E\left[\Phi^{-1}(\nu)\sum_{\iota=1}^{\nu}\boldsymbol{\upsilon}(\iota)\varepsilon_o^*(\iota)\middle|\boldsymbol{\upsilon}(\iota), \iota = 1,2,\mathrm{K}\,,\nu\right]\right] \\
&= \boldsymbol{\omega}_o + E\left[\Phi^{-1}(\nu)\sum_{\iota=1}^{\nu}\boldsymbol{\upsilon}(\iota)E\left[\varepsilon_o^*(\iota)\middle|\boldsymbol{\upsilon}(\iota), \iota = 1,2,\mathrm{K}\,,\nu\right]\right]
\end{aligned}
$$

since both $\mathbf{\Phi}(n)$ and $\mathbf{u}(i)$ are uniquely defined by $\mathbf{u}(i)$ and further conditioning has no effect.

Since $e_o(n)$ is zero mean and independent of $\mathbf{u}(i)$ we have

$$
E[\hat{\mathbf{w}}(n)] = \boldsymbol{\omega}_o + E\left[\Phi^{-1}(\nu)\sum_{\iota=1}^{\nu}\boldsymbol{\upsilon}(\iota)\underbrace{E\left[\varepsilon_o^*(\iota)\middle|\boldsymbol{\upsilon}(\iota), \iota = 1,2,\mathrm{K}\,,\nu\right]}_{0}\right]
$$

or

$$
\boxed{E[\hat{\mathbf{w}}(n)] = \boldsymbol{\omega}_o\,,\ n \geq M}
$$

This suggests that with RLS algorithm, convergent of the mean is achieved for $n \geq M$. We note that we required $n \to \infty$ for LMS convergence of the mean.

Variance of the Misalignment Vector (an interesting result)

We've shown that with RLS we can expect $\hat{\mathbf{w}}(n)$ to converge to $\mathbf{w}_o$. Next we analyze the variance of the weight-error or misalignment vector to determine how far the filter coefficients wander around the optimal solution. The outline for the derivation is as follows:

Goal: Determine $E[\varepsilon^H(n)\varepsilon(n)]$

Step 1: Compute $\varepsilon(n)$

Step 2: Determine $\mathbf{K}(n) = E[\varepsilon(n)\varepsilon^H(n)]$

Step 3: Determine $E[\varepsilon^H(n)\varepsilon(n)] = \mathrm{tr}\{E[\varepsilon^H(n)\varepsilon(n)]\} = \mathrm{tr}[\mathbf{K}(n)]$

Step 1

We again assume $\lambda = 1$ thus

$$
\begin{aligned}
\Phi(n) &= \sum_{i=1}^{n}\mathbf{u}(i)\mathbf{u}^H(i) \\
\mathbf{z}(n) &= \sum_{i=1}^{n}\mathbf{u}(i)d^*(i)
\end{aligned}
.
$$

Substitution into the solution to the normal equations we have

$$
\begin{aligned}
\hat{\mathbf{w}}(n) &= \Phi^{-1}(n)\mathbf{z}(n) \\
&= \mathbf{w}_o + \Phi^{-1}(n)\sum_{i=1}^{n}\mathbf{u}(i)e_o^*(i)
\end{aligned}
$$

or (take this ε to be the misalignment vector--not the LS cost function!)

$$\begin{aligned}\varepsilon(n) &= \hat{\mathbf{w}}(n) - \mathbf{w}_o \\ &= \Phi^{-1}(n)\sum_{i=1}^{n}\mathbf{u}(i)e_o^*(i)\end{aligned}.$$

Step 2

Next, we use the definition of the weight-error correlation matrix and substitute

$$\begin{aligned}\mathbf{K}(n) &= E[\varepsilon(n)\varepsilon^H(n)] \\ &= E\left[\Phi^{-1}(n)\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{u}(i)e_o^*(i)e_o(j)\mathbf{u}^H(j)\Phi^{-H}(n)\right]\end{aligned}.$$

Using our conditional expectation property we have (note that $\Phi^{-H} = \Phi^{-1}$ since $\Phi = \Phi^H$)

$$\mathbf{K}(n) = E\left[\Phi^{-1}(\nu)\sum_{\iota=1}^{\nu}\sum_{\varphi=1}^{\nu}\boldsymbol{\upsilon}(\iota)E[\varepsilon_o^*(\iota)\varepsilon_o(\varphi)]\boldsymbol{\upsilon}^H(\varphi)\Phi^{-1}(\nu)\right].$$

We assume the measurement error, $e_o(n)$ is white with σ^2 variance

$$E[e_o(i)\varepsilon_o^*(\varphi)] = \begin{cases}\sigma^2, & \iota = \varphi \\ 0, & o.\omega.\end{cases}.$$

Therefore

$$\begin{aligned}\mathbf{K}(n) &= \sigma^2 E\left[\Phi^{-1}(\nu)\sum_{\iota=1}^{\nu}\sum_{\varphi=1}^{\nu}\boldsymbol{\upsilon}(\iota)\boldsymbol{\upsilon}^H(\varphi)\Phi^{-1}(\nu)\right] \\ &= \sigma^2 E\left[\Phi^{-1}(\nu)\sum_{\iota=1}^{\nu}\boldsymbol{\upsilon}(\iota)\boldsymbol{\upsilon}^H(\iota)\Phi^{-1}(\nu)\right] \\ &= \sigma^2 E[\Phi^{-1}(\nu)\Phi(\nu)\Phi^{-1}(\nu)] \\ &= \sigma^2 E[\Phi^{-1}(\nu)]\end{aligned}.$$

Next we employ independence assumptions

1) input vectors $\mathbf{u}(1),\boldsymbol{\upsilon}(2),\Lambda\ ,\boldsymbol{\upsilon}(\nu)$ constitute a sequence of statistically independent (iid) vectors.
2) input vectors $\mathbf{u}(1),\boldsymbol{\upsilon}(2),\Lambda\ ,\boldsymbol{\upsilon}(\nu)$ are drawn from a stochastic process with a multivariate Gaussian distribution with zero mean and ensemble-averaged correlation matrix **R**.

Given the above assumptions, it can be shown that

$$E[\Phi^{-1}(\nu)] = \frac{1}{\nu - M - 1}\mathbf{P}^{-1},\ \nu > M + 1$$

thus

$$E[\varepsilon(\nu)\varepsilon^H(\nu)] = \mathbf{K}(\nu)$$
$$= \frac{\sigma^2}{\nu - M - 1}\mathbf{P}^{-1}, \ \nu > M+1 .$$

Step 3

We can compute the variance in the misalignment vector $\boldsymbol{\varepsilon}(n)$ or equivalently the mean-squared error in the weight vector $\hat{\mathbf{w}}(n)$ as follows:

$$\underbrace{E[\varepsilon^H(n)\varepsilon(n)]}_{\text{var of misalignment vector}} = \operatorname{tr}\{E[\varepsilon^H(n)\varepsilon(n)]\}$$
$$= \operatorname{tr}\{E[\varepsilon(n)\varepsilon^H(n)]\}$$
$$= \operatorname{tr}[\mathbf{K}(n)]$$
$$= \frac{\sigma^2}{n - M - 1}\operatorname{tr}[\mathbf{R}^{-1}], \ n > M + 1$$
$$= \frac{\sigma^2}{n - M - 1}\sum_{i=1}^{M}\frac{1}{\lambda_i}$$

where $\lambda_1, ..., \lambda_M$ are the eigenvalues of $\mathbf{R}$.

Observations

1) the variance in the misalignment vector is initially sensitive to the smallest eigenvalue λ_{min}. Therefore unlike LMS in which large eigenvalue spread *slowed* convergence, a single small eigenvalue can *initially* cause RLS to behave badly, i.e. possibly diverge.

2) the variance in the misalignment vector decays almost linearly with n. Thus $\hat{\mathbf{w}}(n)$ converges in the norm to $\mathbf{w}_o$ almost *linearly* with time.