A Derivation of the Steady-State MSE of RLS: Stationary and Nonstationary Cases

Phil Schniter

Nov. 20, 2001*

Abstract

In this report we combine the approach of Yousef and Sayed [1] with that of Rupp and Sayed [2] to derive the steady-state mean-squared error (MSE) of the recursive least squares (RLS) algorithm in both stationary and non-stationary environments. Comparisons with the steady-state MSE of LMS are included.

1 Introduction

The recursive least squares (RLS) parameter update equation can be written as

$$\boldsymbol{w}(n) = \boldsymbol{w}(n-1) + \underbrace{\boldsymbol{P}(n)\boldsymbol{u}(n)}_{\boldsymbol{k}(n)} \underbrace{\big(d(n) - \boldsymbol{w}^H(n-1)\boldsymbol{u}(n)\big)^*}_{\xi^*(n)}$$

where $\boldsymbol{P}(n)$ is the inverse of the deterministic autocorrelation matrix $\boldsymbol{\Phi}(n)$

$$\begin{aligned} \boldsymbol{P}(n) &= \boldsymbol{\Phi}^{-1}(n) \\ \boldsymbol{\Phi}(n) &= \sum_{i=1}^{n} \lambda^{n-i}\boldsymbol{u}(i)\boldsymbol{u}^H(i) + \delta\lambda^n \boldsymbol{I} \end{aligned}$$

with $0 < \lambda \leq 1$. We assume a nonstationary environment with time-varying Wiener filter $\boldsymbol{w}_\star(n)$ that evolves as a random walk:

$$\begin{aligned} d(n) &= \boldsymbol{w}_\star^H(n-1)\boldsymbol{u}(n) + \xi_\star(n) \\ \boldsymbol{w}_\star(n) &= \boldsymbol{w}_\star(n-1) + \boldsymbol{q}(n) \end{aligned} \tag{1}$$

$\xi_\star(n)$ denotes the MMSE version of $\xi(n)$ which, due to the MMSE orthogonality principle, is uncorrelated with $\boldsymbol{u}(n)$:

$$\mathrm{E}\{\boldsymbol{u}(n)\xi_\star^*(n)\} = \boldsymbol{0}$$

*Revised November 14, 2003.

Other statistical assumptions will be stated later.

We will make extensive use of the parameter error vector $\tilde{\boldsymbol{w}}(n)$, the a posteriori error $e_p(n)$, and the a priori error $e_a(n)$.

$$\begin{aligned}
\tilde{\boldsymbol{w}}(n) &:= \boldsymbol{w}_\star(n) - \boldsymbol{w}(n) \\
e_p(n) &:= \big(d(n) - \boldsymbol{w}^H(n)\boldsymbol{u}(n)\big) - \xi_\star(n) \\
&= \big(\tilde{\boldsymbol{w}}^H(n) - \boldsymbol{q}(n)\big)^H \boldsymbol{u}(n) \\
e_a(n) &:= \big(d(n) - \boldsymbol{w}^H(n-1)\boldsymbol{u}(n)\big) - \xi_\star(n) \;=\; \xi(n) - \xi_\star(n) \qquad (2) \\
&= \tilde{\boldsymbol{w}}^H(n-1)\boldsymbol{u}(n)
\end{aligned}$$

2 Energy Relation

In this section we will derive a fundamental deterministic energy relationship.

$$\begin{aligned}
\boldsymbol{w}(n) &= \boldsymbol{w}(n-1) + \boldsymbol{P}(n)\boldsymbol{u}(n)\xi^*(n) \\
\tilde{\boldsymbol{w}}(n) &= \tilde{\boldsymbol{w}}(n-1) - \boldsymbol{P}(n)\boldsymbol{u}(n)\xi^*(n) + \boldsymbol{q}(n) \qquad (3) \\
\underbrace{(\tilde{\boldsymbol{w}}(n) - \boldsymbol{q}(n))^H\boldsymbol{u}(n)}_{e_p(n)} &= \underbrace{\tilde{\boldsymbol{w}}^H(n-1)\boldsymbol{u}(n)}_{e_a(n)} - \underbrace{\boldsymbol{u}^H(n)\boldsymbol{P}(n)\boldsymbol{u}(n)}_{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}}\xi(n) \qquad (4)
\end{aligned}$$

Note the use of the weighted Euclidean norm

$$\|\boldsymbol{a}\|^2_{\boldsymbol{B}} \;:=\; \boldsymbol{a}^H\boldsymbol{B}\boldsymbol{a}$$

where $\boldsymbol{B}$ is a Hermitian positive-semi-definite matrix. Equation (4) implies that

$$\begin{aligned}
\xi(n) &= \|\boldsymbol{u}(n)\|^{-2}_{\boldsymbol{P}(n)}\big(e_a(n) - e_p(n)\big) \qquad \text{when} \qquad \boldsymbol{u}(n) \neq 0 \\
\tilde{\boldsymbol{w}}(n) - \boldsymbol{q}(n) &= \begin{cases} \tilde{\boldsymbol{w}}(n-1) - \|\boldsymbol{u}(n)\|^{-2}_{\boldsymbol{P}(n)}\boldsymbol{P}(n)\boldsymbol{u}(n)\big(e_a(n) - e_p(n)\big)^* & \boldsymbol{u}(n) \neq 0 \\ \tilde{\boldsymbol{w}}(n-1) & \boldsymbol{u}(n) = 0 \end{cases} \\
&= \tilde{\boldsymbol{w}}(n-1) - \bar{\mu}(n)\boldsymbol{P}(n)\boldsymbol{u}(n)\big(e_a(n) - e_p(n)\big)^* \qquad (5)
\end{aligned}$$

where we use the psuedo-inverse to define

$$\bar{\mu}(n) \;:=\; \Big(\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}\Big)^\dagger$$

Taking the $\boldsymbol{P}^{-1}(n)$-weighted norm of both sides of (5), we get $\|\tilde{\boldsymbol{w}}(n) - \boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}$ for the left side and the following quantity for the right.

$$\begin{aligned}
&\Big(\tilde{\boldsymbol{w}}^H(n-1) - \bar{\mu}(n)e_a(n)\boldsymbol{u}^H(n)\boldsymbol{P}(n) + \bar{\mu}(n)e_p(n)\boldsymbol{u}^H(n)\boldsymbol{P}(n)\Big)\boldsymbol{P}^{-1}(n) \\
&\cdot\Big(\tilde{\boldsymbol{w}}(n-1) - \bar{\mu}(n)e_a^*(n)\boldsymbol{P}(n)\boldsymbol{u}(n) + \bar{\mu}(n)e_p^*(n)\boldsymbol{P}(n)\boldsymbol{u}(n)\Big) \\
&\qquad = \|\tilde{\boldsymbol{w}}(n-1)\|^2_{\boldsymbol{P}^{-1}(n)} - \bar{\mu}(n)|e_a(n)|^2 + \bar{\mu}(n)e_p^*(n)e_a(n) \\
&\qquad\quad - \bar{\mu}(n)|e_a(n)|^2 + \bar{\mu}(n)|e_a(n)|^2 - \bar{\mu}(n)e_a(n)e_p^*(n) \\
&\qquad\quad + \bar{\mu}(n)e_p(n)e_a^*(n) - \bar{\mu}(n)e_p(n)e_a^*(n) + \bar{\mu}(n)|e_p(n)|^2 \\
&\qquad = \|\tilde{\boldsymbol{w}}(n-1)\|^2_{\boldsymbol{P}^{-1}(n)} - \bar{\mu}(n)|e_a(n)|^2 + \bar{\mu}(n)|e_p(n)|^2
\end{aligned}$$

The energy relation is then summarized as

$$\|\tilde{\boldsymbol{w}}(n)-\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}+\bar{\mu}(n)|e_a(n)|^2 \;=\; \|\tilde{\boldsymbol{w}}(n-1)\|^2_{\boldsymbol{P}^{-1}(n)}+\bar{\mu}(n)|e_p(n)|^2 \qquad (6)$$

3 Random Walk Assumptions

We assume the following about the random-walk driving-noise $\{\boldsymbol{q}(n)\}$ in (1).

A.1)
$$\begin{aligned} \mathrm{E}\{\boldsymbol{q}(n)\} &= 0 \\ \mathrm{E}\{\boldsymbol{q}(n)\boldsymbol{q}^H(n-k)\} &= \boldsymbol{Q}\delta_k \quad \text{where } \delta_k \text{ denotes the Kronecker delta} \\ \{\boldsymbol{q}(n)\} &\perp\!\!\!\perp \{\boldsymbol{u}(n)\} \\ \{\boldsymbol{q}(n)\} &\perp\!\!\!\perp \{\boldsymbol{\xi}(n)\} \end{aligned}$$

Examining the expectation of the leftmost term of the energy relation (6), we have

$$\begin{aligned} &\mathrm{E}\big\{\|\tilde{\boldsymbol{w}}(n)-\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} \\ &\quad= \mathrm{E}\big\{\|\tilde{\boldsymbol{w}}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} - 2\Re\,\mathrm{E}\{\boldsymbol{q}^H(n)\boldsymbol{P}^{-1}(n)\tilde{\boldsymbol{w}}(n)\big\} + \mathrm{E}\big\{\|\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} \end{aligned} \qquad (7)$$

From (3) we have

$$\tilde{\boldsymbol{w}}(n) \;=\; \tilde{\boldsymbol{w}}(n-1)+\boldsymbol{q}(n)-\boldsymbol{P}(n)\boldsymbol{u}(n)\xi^*(n)$$

allowing us to simplify the second term on the right side of (7):

$$\begin{aligned} &\mathrm{E}\big\{\boldsymbol{q}^H(n)\boldsymbol{P}^{-1}(n)\tilde{\boldsymbol{w}}(n)\big\} \\ &\quad= \mathrm{E}\big\{\boldsymbol{q}^H(n)\boldsymbol{P}^{-1}(n)\tilde{\boldsymbol{w}}(n-1)\big\} + \mathrm{E}\big\{\|\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} - \mathrm{E}\big\{\boldsymbol{q}^H(n)\boldsymbol{u}(n)\xi^*(n)\big\} \\ &\quad= \mathrm{E}\big\{\|\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} \end{aligned}$$

where the first and third terms vanished as a result of A.1). Plugging the previous (real-valued) expression into (7) yields

$$\mathrm{E}\big\{\|\tilde{\boldsymbol{w}}(n)-\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} \;=\; \mathrm{E}\big\{\|\tilde{\boldsymbol{w}}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} - \mathrm{E}\big\{\|\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} \qquad (8)$$

4 Steady-State Analysis

We claim that when the adaptation algorithm has reached "steady state",

$$\mathrm{E}\big\{\|\tilde{\boldsymbol{w}}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} \;=\; \mathrm{E}\big\{\|\tilde{\boldsymbol{w}}(n-1)\|^2_{\boldsymbol{P}^{-1}(n)}\big\}$$

Taking the expectation of (6) and incorporating (8), we find the following steady state relationship:

$$\mathrm{E}\big\{\bar{\mu}(n)|e_a(n)|^2\big\} \;=\; \mathrm{E}\big\{\bar{\mu}(n)|e_p(n)|^2\big\} + \mathrm{E}\big\{\|\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} \qquad (9)$$

To proceed further, we make a few assumptions:

A.2) $\mathrm{E}\{\boldsymbol{u}(n)\boldsymbol{u}^H(n)\} = \boldsymbol{R}$ and $\boldsymbol{R} > 0$

A.3) $\{\xi_\star(n)\}$ i.i.d. $\quad\Rightarrow\quad \{\xi_\star(n)\} \perp\!\!\!\perp \{\tilde{\boldsymbol{w}}(n-1)\}$
$\{\xi_\star(n)\} \perp\!\!\!\perp \{\boldsymbol{u}(n)\} \qquad\qquad \{\xi_\star(n)\} \perp\!\!\!\perp \{e_a(n)\}$

A.4) $\mathrm{E}\{\boldsymbol{P}(n)\boldsymbol{u}(n)\boldsymbol{u}^H(n)\} = \mathrm{E}\{\boldsymbol{P}(n)\}\,\mathrm{E}\{\boldsymbol{u}(n)\boldsymbol{u}^H(n)\}$ (e.g., from $\lambda \approx 1$)

A.5) $\mathrm{E}\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}|e_a(n)|^2\} = \mathrm{E}\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}\}\,\mathrm{E}\{|e_a(n)|^2\}$ (e.g., from large M)

A.6) $\lim_{n\to\infty}\mathrm{E}\{\boldsymbol{\Phi}^{-1}(n)\} = \left(\lim_{n\to\infty}\mathrm{E}\{\boldsymbol{\Phi}(n)\}\right)^{-1}$ (e.g., Wishart theory [3, p. 452])

First, using the implications of A.3), we have

$$\begin{aligned}
J(n) &:= \mathrm{E}\{|\xi(n)|^2\} \\
&= \mathrm{E}\{|e_a(n)+\xi_\star(n)|^2\} \\
&= \mathrm{E}\{|e_a(n)|^2\} + \mathrm{E}\{|\xi_\star(n)|^2\} \\
&= \mathrm{E}\{|e_a(n)|^2\} + J_{\min} \\
J_{\text{emse}} &:= \lim_{n\to\infty} J(n) - J_{\min} \\
&= \lim_{n\to\infty} \mathrm{E}\{|e_a(n)|^2\}
\end{aligned}$$

Then, from (2), (4), and A.3), we see that the first term on the right side of (9) can be written as

$$\begin{aligned}
&\mathrm{E}\{\bar{\mu}(n)|e_p(n)|^2\} \\
&= \mathrm{E}\left\{\bar{\mu}(n)\left|e_a(n) - \|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}\xi(n)\right|^2\right\} \\
&= \mathrm{E}\left\{\bar{\mu}(n)\left|\left(1-\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}\right)e_a(n) + \|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}\xi_\star(n)\right|^2\right\} \\
&= \mathrm{E}\left\{\bar{\mu}(n)\left|e_a(n)\right|^2\right\} - 2\,\mathrm{E}\left\{|e_a(n)|^2\right\} + \mathrm{E}\left\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}|e_a(n)|^2\right\} + \mathrm{E}\left\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}|\xi_\star(n)|^2\right\}
\end{aligned}$$

So the steady-state relationship (9) becomes (after cancellation of $\mathrm{E}\{\bar{\mu}(n)|e_a(n)|^2\}$ terms)

$$2\,\mathrm{E}\{|e_a(n)|^2\} = \mathrm{E}\left\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}|e_a(n)|^2\right\} + \mathrm{E}\left\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}|\xi_\star(n)|^2\right\} + \mathrm{E}\{\|\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\} \quad (10)$$

Looking at the first term on the right side of (10), we have, from A.4) and A.5),

$$\begin{aligned}
\mathrm{E}\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}|e_a(n)|^2\} &= \mathrm{E}\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)}\}\,\mathrm{E}\{|e_a(n)|^2\} \\
&= \mathrm{tr}\Big(\mathrm{E}\{\boldsymbol{P}(n)\boldsymbol{u}(n)\boldsymbol{u}^H(n)\}\Big)\,\mathrm{E}\{|e_a(n)|^2\} \\
&= \mathrm{tr}\Big(\mathrm{E}\{\boldsymbol{P}(n)\}\,\mathrm{E}\{\boldsymbol{u}(n)\boldsymbol{u}^H(n)\}\Big)\,\mathrm{E}\{|e_a(n)|^2\} \\
&= \mathrm{tr}\Big(\mathrm{E}\{\boldsymbol{P}(n)\}\boldsymbol{R}\Big)\,\mathrm{E}\{|e_a(n)|^2\}
\end{aligned}$$

where $\mathrm{E}\big\{\boldsymbol{P}(n)\big\}$ requires further investigation. Using the fact that

$$\begin{aligned}\mathrm{E}\big\{\boldsymbol{\Phi}(n)\big\} &= \sum_{i=1}^{n} \lambda^{n-i}\,\mathrm{E}\{\boldsymbol{u}(n)\boldsymbol{u}^H(n)\} + \delta\lambda^n \boldsymbol{I} \\ &= \begin{cases} \frac{1-\lambda^n}{1-\lambda}\boldsymbol{R} + \delta\lambda^n\boldsymbol{I} & \lambda < 1 \\ n\boldsymbol{R} + \delta\boldsymbol{I} & \lambda = 1 \end{cases}\end{aligned}$$

we find, via A.6),

$$\begin{aligned}\lim_{n\to\infty} \mathrm{E}\big\{\boldsymbol{P}(n)\big\} &= \Big(\lim_{n\to\infty} \mathrm{E}\big\{\boldsymbol{\Phi}(n)\big\}\Big)^{-1} \\ &= \begin{cases} (1-\lambda)\boldsymbol{R}^{-1} & \lambda < 1 \\ 0 & \lambda = 1 \end{cases}\end{aligned}$$

implying that

$$\lim_{n\to\infty} \mathrm{E}\Big\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)} |e_a(n)|^2\Big\} = \begin{cases} M(1-\lambda)J_{\mathrm{emse}} & \lambda < 1 \\ 0 & \lambda = 1 \end{cases}$$

Looking next at the second term on the right side of (10), we get from A.3) and A.6)

$$\lim_{n\to\infty} \mathrm{E}\Big\{\|\boldsymbol{u}(n)\|^2_{\boldsymbol{P}(n)} |\xi_\star(n)|^2\Big\} = \begin{cases} M(1-\lambda)J_{\min} & \lambda < 1 \\ 0 & \lambda = 1 \end{cases}$$

For the last term on the right side of (10), assumptions A.1) and A.2) yield

$$\begin{aligned}\mathrm{E}\big\{\|\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} &= \mathrm{E}\big\{\boldsymbol{q}^H(n)\boldsymbol{\Phi}(n)\boldsymbol{q}(n)\big\} \\ &= \mathrm{tr}\,\big(\mathrm{E}\big\{\boldsymbol{q}(n)\boldsymbol{q}^H(n)\boldsymbol{\Phi}(n)\big\}\big) \\ &= \mathrm{tr}\left(\mathrm{E}\big\{\boldsymbol{q}(n)\boldsymbol{q}^H(n)\big\}\sum_{i=1}^{n}\lambda^{n-i}\,\mathrm{E}\big\{\boldsymbol{u}(n)\boldsymbol{u}^H(n)\big\} + \delta\lambda^n E\big\{\boldsymbol{q}(n)\boldsymbol{q}^H(n)\big\}\right) \\ &= \mathrm{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big)\sum_{i=1}^{n}\lambda^{n-i} + \delta\lambda^n\,\mathrm{tr}\big(\boldsymbol{Q}\big) \\ &= \begin{cases} \frac{1-\lambda^n}{1-\lambda}\,\mathrm{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big) + \delta\lambda^n\,\mathrm{tr}\big(\boldsymbol{Q}\big) & \lambda < 1 \\ n\,\mathrm{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big) + \delta\,\mathrm{tr}\big(\boldsymbol{Q}\big) & \lambda = 1 \end{cases}\end{aligned}$$

so that

$$\lim_{n\to\infty} \mathrm{E}\big\{\|\boldsymbol{q}(n)\|^2_{\boldsymbol{P}^{-1}(n)}\big\} = \begin{cases} (1-\lambda)^{-1}\,\mathrm{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big) & \lambda < 1 \\ \infty & \lambda = 1 \text{ and } \boldsymbol{Q} \neq \boldsymbol{0} \\ 0 & \lambda = 1 \text{ and } \boldsymbol{Q} = \boldsymbol{0} \end{cases}$$

Collecting our findings on the asymptotic behavior of (10), we have

$$2J_{\mathrm{emse}} = \begin{cases} M(1-\lambda)J_{\mathrm{emse}} + M(1-\lambda)J_{\min} + (1-\lambda)^{-1}\,\mathrm{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big) & \lambda < 1 \\ \infty & \lambda = 1 \text{ and } \boldsymbol{Q} \neq \boldsymbol{0} \\ 0 & \lambda = 1 \text{ and } \boldsymbol{Q} = \boldsymbol{0} \end{cases}$$

which, for the $\lambda < 1$ case, gives

$$J_{\text{emse, RLS}} = \frac{M(1-\lambda)J_{\min} + (1-\lambda)^{-1}\operatorname{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big)}{2 - M(1-\lambda)} \quad (11)$$

$$\approx \frac{1}{2}\Big(M(1-\lambda)J_{\min} + (1-\lambda)^{-1}\operatorname{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big)\Big) \quad \text{when} \quad |M(1-\lambda)| \ll 2 \quad (12)$$

Note the EMSE component due to $J_{\min}$ and due to time variations (i.e., $\boldsymbol{Q}$).

5 Optimum Forgetting Factor

The value of λ that minimizes J_{emse} can be obtained by zeroing the partial derivative of (12) with respect to $(1-\lambda)$. This yields

$$\lambda_{\text{opt}} \approx 1 - \sqrt{\frac{\operatorname{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big)}{M J_{\min}}} \quad (13)$$

$$J_{\text{emse, RLS}}\big|_{\text{opt}} \approx \sqrt{M J_{\min} \operatorname{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big)} \quad (14)$$

6 Comparison to LMS

Recall the counterparts of (11)–(14) for LMS:

$$J_{\text{emse, LMS}} = \frac{\mu \operatorname{tr}(\boldsymbol{R}) J_{\min} + \mu^{-1}\operatorname{tr}(\boldsymbol{Q})}{2 - \mu \operatorname{tr}(\boldsymbol{R})} \quad (15)$$

$$\approx \frac{1}{2}\Big(\mu \operatorname{tr}(\boldsymbol{R}) J_{\min} + \mu^{-1}\operatorname{tr}(\boldsymbol{Q})\Big) \quad \text{when} \quad \mu \operatorname{tr}(\boldsymbol{R}) \ll 2 \quad (16)$$

$$\mu_{\text{opt}} \approx \sqrt{\frac{\operatorname{tr}(\boldsymbol{Q})}{J_{\min}\operatorname{tr}(\boldsymbol{R})}} \quad (17)$$

$$J_{\text{emse, LMS}}\big|_{\text{opt}} \approx \sqrt{J_{\min}\operatorname{tr}(\boldsymbol{R})\operatorname{tr}(\boldsymbol{Q})} \quad (18)$$

Comparing the optimal values of EMSE for LMS and RLS we find

$$\frac{J_{\text{emse, LMS}}\big|_{\text{opt}}}{J_{\text{emse, RLS}}\big|_{\text{opt}}} = \sqrt{\frac{\operatorname{tr}(\boldsymbol{R})\operatorname{tr}(\boldsymbol{Q})}{M \operatorname{tr}\big(\boldsymbol{Q}\boldsymbol{R}\big)}} \quad (19)$$

Thus the relationship between $\boldsymbol{R}$ and $\boldsymbol{Q}$ will determine which of the two algorithms gives lower steady-state MSE. A few instructive cases are presented below [3]. It is interesting to note that the simple LMS algorithm can have exhibit *tracking* performance superior to the computationally-intensive RLS algorithm in certain environments. However, it should be noted that the *convergence* of RLS is typically much faster than that of LMS.

- $\underline{\boldsymbol{Q} = \sigma_q^2 \boldsymbol{I}}$: In this case we find that the two algorithms provide essentially the same level of steady-state MSE.

$$J_{\text{emse}} \approx \sqrt{M J_{\min} \sigma_q^2 \operatorname{tr}(\boldsymbol{R})}$$

- $\underline{\boldsymbol{Q} = c\boldsymbol{R}}$: In this case LMS yields lower steady-state MSE:

$$\frac{J_{\text{emse, LMS}}\big|_{\text{opt}}}{J_{\text{emse, RLS}}\big|_{\text{opt}}} \;=\; \sqrt{\frac{\operatorname{tr}(\boldsymbol{R})^2}{M \operatorname{tr}(\boldsymbol{R}^2)}} \;<\; 1$$

- $\underline{\boldsymbol{Q} = c\boldsymbol{R}^{-1}}$: In this case RLS yields lower steady-state MSE:

$$\frac{J_{\text{emse, LMS}}\big|_{\text{opt}}}{J_{\text{emse, RLS}}\big|_{\text{opt}}} \;=\; \frac{1}{M}\sqrt{\operatorname{tr}(\boldsymbol{R}) \operatorname{tr}(\boldsymbol{R}^{-1})} \;>\; 1$$

The inequalities above can be proven using the Cauchy-Schwarz inequality for vectors: $|\boldsymbol{x}^H \boldsymbol{y}|^2 \leq \|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2$.

References

[1] N. B. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Trans. on Signal Processing*, vol. 49, pp. 314–324, Feb. 2001.

[2] M. Rupp and A. H. Sayed, "Robustness of Gauss-Newton recursive methods: A deterministic feedback analysis," *Signal Processing*, vol. 50, pp. 165–187, 1996.

[3] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 4th ed., 2001.